



THE COMMUNITY OF PRACTICE ON
CROP MODELING



CoP on Crop Modeling Mini-Grants Report 2018

Molero Milan, Anabel; Reynolds, Matthew; Asseng, Senthold; Kholova, Jana; Ramirez-Villegas, Julian; Krupnik, Timothy J.; Jones, Peter. (2018)

CGIAR Platform for Big Data in Agriculture



CoP on Crop Modeling Mini-Grants Report 2018

Introduction

The Community of Practice on Crop Modeling (CoP CM) started a **mini-grant program** in 2017 to provide our CoP members from CGIAR and external partners with mini-grants (10K-20K) to **facilitate the development of key activities, tools, datasets, and model analysis** that can facilitate CGIAR's crop modeling research and could achieve some boost/impact and promote collaborations between CGIAR centers and partners. The results of the funded projects will be shared among the CoP via webinars and news feeds.

The first call was sent out on November 2nd 2017 within the CoP on CM members and 5 competitive proposals were received and evaluated. 3 out of the 5 proposals were selected and have been developing their activities during 2018:

- 1) **CGIAR Data Search for Crop Modeling** | Senthold Asseng, UF
- 2) **Development of system modeling platform to guide agro-ecosystem specific interventions to enhance post-rainy sorghum production in India** | Jana Kholova, ICRISAT
- 3) **Determination of bean breeding targets in Colombia** | Julian Ramirez-Villegas, CIAT

During 2018, due to funds availability, three extra projects were awarded with the CoP on Crop Modeling mini-grant program, one from the 2017 call, one from the GeoSpatial CoP call, and one from a new project initiated within the CoP for analyzing the massive International Wheat Improvement Network (IWIN) dataset to make the database more interoperable and accessible:

- 4) **Combining crop and disease modeling with numerical weather forecasting to inform wheat blast early warning systems in Bangladesh, Brazil and beyond** | Timothy J. Krupnik, CIMMYT



THE COMMUNITY OF PRACTICE ON
CROP MODELING



- 5) **Gridded climate datasets and tools for CGIAR modelling applications** | Peter Jones, Waen Associates, Ltd.
- 6) **Better Insights into Germplasm through IWIN (BIGIWIN)** | Urs Schulthess and Wei Xiong, CIMMYT

#	Title of Project	Project leader/s, Affiliation	Lead Partner	Starting Date	End Date	Main Outputs	Status	Amount funded (\$)
1	CGIAR Data Search for Crop Modeling	Senthold Asseng, University of Florida (UF) Gerrit Hoogenboom, UF Cheryl Porter, UF	Matthew Reynolds, CIMMYT Diego Pequeno, CIMMYT	1 st Jan 2018	30 th June 2018	<ul style="list-style-type: none"> - Evaluation of GARDIAN as search engine - Conversion of 2 CGIAR field datasets into model-ready data - Flexible format data translation tool 	Finished June 2018	20 K
2	Development of system modeling platform to guide agro-ecosystem specific interventions to enhance post-rainy sorghum production in India	Jana Kholova, ICRISAT	Swarna Ronanki, Indian Institute of Millet Research	1 st Jan 2018	31 st Dec 2021	<ul style="list-style-type: none"> - Sorghum simulation grid refined and validated - Use of modeling as a decision-making breeding support tool to design location-specific fine-tuning (GxM) agronomic interventions - In-vivo multi-location trials of relevant GxM predicted options 	On-going	20 K
3	Determination of bean breeding targets in Colombia	Julian Ramirez-Villegas, CIAT	Andy Challinor, University of Leeds, UK	1 st April 2018	31 st March 2019	<ul style="list-style-type: none"> - Calibrated bean model for Colombian conditions - Identify stresses and selection sites for breeding bean for abiotic stress in 	On-going	20 K

						Colombia using crop climate modelling - Updated breeding strategy based on modeling results to help target activities to specific environments		
4	Combining crop and disease modeling with numerical weather forecasting to inform wheat blast early warning systems in Bangladesh, Brazil and beyond	Timothy J. Krupnik, CIMMYT	Jose Mauricio Fernandes, U. de Passo Fundo, Brazil	1 st July 2018	30 th Sept 2019	- Full integration of crop models into the MoT EWS – made available online - Outbreak risk maps in Bangladesh and Brazil	On-going	20 K
5	Gridded climate datasets and tools for CGIAR modelling applications	Peter Jones, Waen Associates, Ltd.	N/A	1 st July 2018	30 th Sept 2018*	- Global climate datasets at the same four resolutions as WorldClim2, organized as direct-access binary files with crop-model-related variables stored in each record by location - Updated MarkSimGCM software, both the web-based and the stand-alone versions, that utilises WorldClim2.	Extended until Feb 19	10 K
6	Better Insights into Germplasm through IWIN (BIGIWIN)	Urs Schulthess, CIMMYT Wei Xiong, CIMMYT	M. Reynolds, T. Payne, J. Crossa, K. Sonder, CIMMYT	1 st Feb 2019	31 st Jan 2021	- Quality controlled, gapfilled and curated IWIN data set with key phenology parameters uploaded to GARDIAN	Starts Feb 2019	20 K



In the following sections, each of the projects will be described and the results and main outputs generated during 2018 presented.

1. CGIAR Data Search for Crop Modeling

Senthold Asseng, Co-Leader of AgMIP-Wheat, University of Florida, USA

Gerrit Hoogenboom, University of Florida, USA

Ms. Cheryl Porter, University of Florida, USA

Chris Villalobos, University of Florida, USA

1.a. Project description

The CGIAR has a vast asset of field experimental data that could be extremely useful for crop model evaluation and improvement. In return, well-tested crop models can guide research in the CGIAR, particularly on adapting germplasm to future climate change and on developing crop management practices for sustainable and environmentally friendly food production. Under the “Organize” component of the Platform for Big Data in Agriculture, the [GARDIAN](#) tool is being developed to provide access to all the data that have been collected at the CGIAR Centers since their inception. The main goal of this project, led by Senthold Asseng from the University of Florida in partnership with Matthew Reynolds from CIMMYT, is to employ the current stage of the search engine [GARDIAN](#) to search, retrieve and translate the data that have been collected at the CGIAR Centers since their inception into AgMIP format field experimental data sets and then made available to crop modelers through the [AgMIP](#) network. The specific objective is to develop an end-to-end case to show that the data is Findable, Accessible, Interoperable and Re-usable (FAIR). As a case study, the CIMMYT wheat database for field data on germplasm with increased heat stress tolerance was explored. To be useful for crop modeling, such field data need to include information on crop management (sowing, N



fertilizer, irrigation, any effect from pest and diseases, lodging), soil (texture, soil organic carbon, bulk density, any constraints to root growth, potential rooting depth, and initial conditions), germplasm characteristics (phenology and yield characteristics, and specific traits, especially those related to heat tolerance), weather (daily max and min temperature, rainfall, solar radiation) and crop measurements (yield, yield components, phenology stages, growth patterns, soil water dynamics). This project started in January 2018 and finished in June 2018.

1.b. Results

This project highlights three very different types of datasets and the steps involved to locate, access, and prepare the data for crop models using AgMIP data translation tools. Three datasets were retrieved from GARDIAN and evaluated for usefulness in crop modeling applications.

The field data sets have been made available to the international crop modeling community through upload to the AgMIP Crop Site Database.

- International Heat Stress Genotype Wheat Experiment:

<https://api.agmip.org/cropsitedb/2/dataset/24c86f43-8b79-4d7e-b10c-f232cef05d66>

- ICARDA Calibrated Wheat:

<https://api.agmip.org/cropsitedb/2/dataset/d127429e-c915-41f1-94cf-824224e60582>

The usefulness of GARDIAN as a user-friendly method of discovering data was also validated. According to the collaborators, the current GARDIAN system is easy to navigate and to perform simple queries on data and publications; however, there is a challenge to making these data useable on a large scale. Currently datasets are stored in



many formats using vocabularies to describe dataset content which are determined on an ad hoc basis by each researcher. When the ICASA Data Dictionary is used as the definition of terms in the dataset, existing AgMIP data translation tools allow rapid translation of data to crop model-specific formats for multiple crop models. Making data useful on a large scale would require annotation of each dataset with terms and definitions that are in alignment with ICASA terms.

A new application was developed that allows Excel spreadsheets to be read directly and relational connections be automatically discovered by the application. This allows the data manager much greater flexibility in organization of the data and also removes the burden of creating csv files from the Excel workbook prior to translation. There were not sufficient funds in this mini-grant to fully incorporate the tool into a user interface, so it runs from a command line only and has limited diagnostics for troubleshooting. The collaborators are currently working with the Module 1 from Big Data Platform in Agriculture to make this translator more generally useful for translating data from spreadsheets including a robust user interface.

The impact of the proposed project is awareness to the crop modeling community of the CGIAR Big Data Platform and the FAIR principles of CGIAR data, and the broader support and impact of the CGIAR to science and policy.

The complete final report that contains the results of the project can be found in [Annex 1](#).

1.c. Outputs

As project outputs we can highlighted:

- 2 Model-ready CGIAR datasets have been made available to the international crop modeling community:



THE COMMUNITY OF PRACTICE ON **CROP MODELING**



- [International Heat Stress Genotype Wheat Experiment](#)
 - [ICARDA Calibrated Wheat](#)
- Blog Post containing the project outcomes in the Crop Modeling CoP website:



GARDIAN and facilitating data interoperability at CGIAR

There is a major gap between the potential value of data collected in agricultural experiments and the value currently obtained through the use of those data. [READ...](#)

2. Development of system modeling platform to guide agro-ecosystem specific interventions to enhance post-rainy sorghum production in India

Jana Kholova, ICRISAT, India.

2.a. Project description

In India, sorghum supports the livelihood of millions food and income-insecure households. Low input agronomic practice and droughts caused by climatic variability are the main reasons why the farmer's yields across post-rainy (rabi) sorghum tract stagnate at ~800kg/ha despite yield potential of same material can attain ~3500kg/ha. With this project, the GEMS team (www.gems.icrisat.org) from ICRISAT and national collaborators from the Indian Institute of Millet Research (IIMR) aim to develop necessary tools to accelerate production of post-rainy (rabi) sorghum yields in India. This will involve using APSIM_sorghum as a decision support tool to design suitable crop and management



THE COMMUNITY OF PRACTICE ON **CROP MODELING**



interventions for particular locations and to optimize rabi sorghum systems productivity. The validated modelling set-up from prior work is being used and being spatially refined to suit this particular purpose. The model refinement includes parameterization of the elite sorghum material into APSIM. This tool will allow assessment of particular GxM fitness to particular production zones (E), development of novel approaches, and its quantitative advantage over standard Maldandi cultivar (M35-1) and standard management practices. Such improved set-up will be further used to design the site-specific optimal genotype x management (GxM) options, which have higher probability to increase productivity in specific regions of rabi-sorghum production tract in India. The most relevant of predicted GxM options will be tested in multi-location trials in-vivo to clearly demonstrate and validate the tool for practical on-ground use of such modeling framework. The project is planned for three years and all the data obtained from the project and resulting publications will be available open-source.

2.b. Partial results

During the first year of the project, Sorghum module of APSIM platform (v7.6) was used to simulate the yields of rabi sorghum in different Indian districts with two different nitrogen applications. Grain and stover yields generated under on-station N management were compared to the farmer practice and the frequencies of yield advantage of on-station over FP were calculated for all simulation units. The simulation experiments showed that the on-station (OS) practice enhanced stover production across central India, but in low yielding environments average yield was 1500 kg/ha with significant grain yield loss. This was on expected lines. Well-fertilized crops establish larger canopy earlier in the season which reflects in higher crop demand for transpiration and leads to earlier water depletion



from the soil. Therefore, on-station crops face water stress earlier in the season with less moisture available to facilitate grain-filling processes.

The study showed that out of 83 investigated districts, only 19 would benefit from OS-N practice in terms of sorghum production. These 19 districts generally encompass regions with deep vertisols with high soil water holding capacity; higher chance of in-crop rainfall – i.e. less chance of severe droughts. In other words, majority of post-rainy sorghum production areas will suffer the grain production losses if higher doses of N fertilizers were applied.

Post rainy Sorghum grain and stover yield response simulation was evaluated using 3 different plant densities:

- Recommended Plant density (RP): 12 plants per m²
- Low density planting (LDP): 6 plants per m²
- High density planting (HDP): 18 plants per m².

The simulation results indicated that there was an increased grain yield with adoption of 6 plants per m² in most of the districts over 12 plants per m². Adoption of high density planting with 18 plants per m² resulted in decreased yields and even complete crop failure in most of the rabi sorghum areas. The beneficial effect of increased plant density was realized in good environmental and soil conditions. Contrary to grain yield, the stover yield increased with increasing plant density from 12 to 18 plants per m² and decreased with reducing the plant density from 12 to 6 plants per m².

During this period, Field experiments were conducted at ICRISAT and ICAR- IIMR during rabi 2017-18 for collecting the data for parameterization of rabi sorghum cultivars into APSIM and for validation of APSIM in split plot design in well watered and water stress conditions at two plant densities and two nitrogen levels.



Grain yields and biomass were compared by running model with observed weather data versus running model with gridded data of NASA and the collaborators observed that simulated yields (both grain yield and biomass) with NASA data were in better agreement with simulated yields with observed weather data. The full project update can be found in Annex 2: Partial Report 2018.

3. Determination of bean breeding targets in Colombia

Julián Ramírez-Villegas, CIAT, Colombia

3.a. Project description

Currently, climate variability and extremes, and particularly drought and heat stress affect crop production across the developing world. Crop improvement plays a central role in enhancing the resilience of the food system to climate variability and climate change. However, the breeding, delivery and adoption of new varieties is constrained by many factors, most notably, the phenotyping capacity and precision and the suitability, frequency and reliability of conditions for selection. Especially for heat stress, no actual targeting for ongoing breeding activities has been conducted. The aim of this project, led by Julian Ramirez-Villegas from CIAT in partnership with Andy Challinor, from the University of Leeds, is to identify heat stress environments for bean breeding in Colombia using crop-climate modeling. More specifically, they will implement the [Target Population of Environments \(TPEs\) methodology](#) for key vulnerable areas in Colombia. The TPE analysis is supported by modeling abiotic stress under current and future climates, itself underpinned by data held at CIAT. Outputs of the project include a breeding strategy that will ultimately make the CIAT bean breeding program more targeted and efficient for abiotic stress tolerance, which will provide wider societal benefits for the Colombian



and other agricultural economies. Success in this project will provide an exemplary case of how breeders and modellers can develop targeted work within one of the Modelling CoP areas, thus inspiring other teams to do the same. The project started in April 2018 and is planned for one year.

3.b. *Partial results*

The CSM-CROPGRO-DRYBEAN model was calibrated and evaluated using detailed growth and development data from an experiment conducted in 2014 in the locality of Villanueva in the Santander department in Colombia. The model was calibrated and evaluated for two varieties included in the trial, a representative commercial check (Calima) and a promising line for drought and heat stress for Colombia (SAB686). The SAB686 variety has been calibrated using the genetic Algorithm (GA) methodology, and Sum square error (SSE) and Distances as function to minimize. A preliminary coefficients set were got for SAB686 and some sensitivity analysis, calibration and evaluation have been performed. The field experiment included seven treatments (i.e. seven different planting dates), out of which two were used for model calibration, and the remaining four were used for model evaluation. The model was calibrated and evaluated for both varieties. In general, the simulation of phenology is found to be accurate for both calibration and evaluation experiments, suggesting the model represents well the phenology of these genotypes. The model was able to simulate well some treatments, however other were poorly simulated, suggesting that there are inherent limitations in the model that require more careful examination and improvement, especially under stress situations and for genotype Calima. Despite these limitations, the model performs sufficiently well for phenology, growth and yield, especially for genotype SAB686.

The analysis of temperature TPEs (Target Population of Environments) indicates distinct responses for the different temperature environments analyzed. Since the future projected



yield of the optimal TPE becomes increasingly similar to that of the warm TPE, it is possible that current optimal growing areas in Colombia are likely to become heat stressed, thus requiring focusing breeding efforts on heat stress adaptation. Ramirez and collaborators are currently reviewing inputs for the model in order to improve some variables (weather, soil, and experimental information) that could contribute better fit for the model and improve the validation. The full project update can be found in Annex 3

3.c. *Outputs from 2018*

As an output from 2018, we can highlight the publication of a Blog post of the project in the Crop Modeling CoP website:



Here's how to do bean breeding the climate-smart way

Projections about the future indicate many parts of the world will see greater heat stress — higher temperatures that can result in more frequent, longer periods of excessive heat for crops. [READ...](#)

4. Combining crop and disease modeling with numerical weather forecasting to inform wheat blast early warning systems in Bangladesh, Brazil and beyond

Timothy J. Krupnik, CIMMYT, Bangladesh

4.a. *Project description*



THE COMMUNITY OF PRACTICE ON **CROP MODELING**



Wheat blast, caused by the fungus *Magnaporthe oryzae* pathotype *Triticum* (MoT), is one of the most problematic wheat diseases globally, and has spread across South America and is now found in South Asia, resulting in significant yield losses undermining food security. Wheat blast infections vary with prevailing climatic conditions, the degree of susceptibility of host cultivars, and the location of infection. Most programs working to mitigate the threat of disease infection focus on host resistance and/or calendar based fungicide application. The latter is a particular concern where environmental and human health is a concern; fungicides are also costly and smallholder wheat farmers may not be able to afford regular calendar based application. An alternative approach is to use weather forecasts and disease models when and where disease might occur, with preventative management advisories delivered to farmers several days in advance of potential disease outbreaks. This project, led by Timothy J. Krupnik from CIMMYT Bangladesh in partnership with Jose Mauricio Fernandes, from the University of Passo Fundo in Brazil, aims to develop such an early warning system (EWS) for wheat blast outbreaks applicable to Bangladesh, Brazil, and beyond using linked crop and disease models and numerical weather forecasting. Because wheat is most susceptible to MoT near flowering, accurate phenological prediction is crucial. EWSs should therefore support estimates for flowering dates in order to increase the risk-prediction accuracy of the model. This project will increase flowering predictability by integrating the Decision Support System for Technology Transfer (DSSAT) model to the MoT EWS framework for Bangladesh and Brazil, thereby allowing improved risk assessments. This work supplements the CIMMYT-led 'Climate Services for Resilient Development (CSRD)' and 'Mitigate Wheat Blast' projects (both USAID funded) to adapt an existing wheat blast forecasting model to South Asian climatic conditions, and to train extension services in its use. This activity also provides the basis for further regional modeling of wheat blast risks using historical climate data and climate change models combined with scenario analysis. These activities



THE COMMUNITY OF PRACTICE ON **CROP MODELING**



will result in both decision support tools and advisories to assist in rational and integrated disease management in South Asia. The project started in July 2018 and will finish by the end of 2019.

4.b. Partial results

During 2018, efforts have been focused to couple a weather-data driven wheat blast development model with the DSSAT-Nwheat model to simulate yield reduction as a function of disease severity in Bangladesh. This work to couple the models is novel and is intended to predict yield in the presence or absence of wheat blast. The original weather-data driven disease model was developed in Brazil and validated for locations where more than a decade of wheat blast observations were available. This model was subsequently transferred to Bangladesh. Disease and crop model coupling will eventually permit model users to explore a range of sowing dates, weather conditions, and cultivar effects on wheat blast infection. Field experiments to calibrate model performance were carried out in Bangladesh in the 2017/18 wheat growing season with six cultivars and five sowing dates. Krupnik and collaborators have been working in actively assembling data from experiments and collecting all genetic coefficients of wheat varieties to feed into modeling efforts.

According to the first experimental results, a general decline in grain yield was observed as sowing date was delayed from November 25th through January 1 that can be associated to terminal heat stress. Preliminary simulated wheat yield and 1,000 grain weight results however showed variable fit with observed data. Efforts to further improve model calibration and fit with observed data will be undertaken in the initial months of 2019. Simulated outputs for virtual experiments conducted using weather data for the 2015/16 wheat season showed a grain and biomass yield decline to model runs with progressively later sowing dates without disease. Simulations of blast disease during the 2016/17 wheat



season reduced yield only marginally, e.g. from 0-3% relative to disease-free simulations, with a slight increase in disease inflicted yield loss with later sowing dates. The results herein presented are partial results and more work is in progress to validate the results. The full project update can be found in Annex 4.

4.c. *Outputs from 2018*

As an output from 2018, we can highlight the publication of a Blog post of the project in the Crop Modeling CoP website:

Cross-continental disease and crop modeling collaborations to beat back wheat blast

Cross-continental collaborations facilitated by the CGIAR Platform for Big Data in Agriculture thrive to beat back the threat of wheat blast in Brazil and Bangladesh.

[READ...](#)



5. Gridded climate datasets and tools for CGIAR modelling applications

Peter Jones, Waen Associates Ltd., UK.

5.a. *Project description*

Agricultural modelling of crops, livestock and households is widely carried out in CGIAR, for a wide variety of purposes, including evaluating different options in time and space in a way that can add massive value to the field-based activities that CGIAR centres undertake. WorldClim v1 has been used very widely in CGIAR's (and other organisations') modelling efforts, as has MarkSim, software that basically allows the user to move from climate data for any location to weather data for the same location that are



characteristic of the particular climate, including future climates as generated by successive generations of IPCC climate modelling activities. Now that WorldClim has been updated with new data, this is an excellent time to update the various MarkSimGCM tools that can be used by the general user, and to provide a set of climate grids that can be used directly by modellers who do their own programming. The proposal will build on the recently-released version 2 of WorldClim, which updates the previous version of the dataset and corrects several errors. WorldClim2 provides gridded climate data by month for the period 1970-2000 for several variables (max and min temp, average temp, precipitation, solar radiation, wind speed and water vapor pressure) at four resolutions (10, 5 and 2.5 minutes and 30 seconds). This project will (1) Increase the functionality of WorldClim2 by developing grid files that include variables by location, as well as providing the basis for deriving a key variable not currently included, number of rain days per month; (2) For one kind of user, reformat WorldClim data into a format that is of direct use to crop and other modellers; (3) For users who do not do their own programming, upgrade MarkSimGCM to work from the WorldClim2 data set, providing weather files that can be used directly in crop modelling software such as DSSAT and APSIM. The project started in July 2018 and due to some technical problems was extended until February 2019.

5.b. Results

This work has updated the MarkSim application data with the new data from WorldClim v2.0, Fick & Hijmans ([2017](#)). The outputs are a new set of MetGrid files at 10, 5 and 2.5 arc minutes and 30 arc seconds, and sets of CLI files, which can be used as input to the stand-alone version of MarkSim or as input to other weather generators. Until now, MetGrid files have not been available to the general user; this changes with this version and they will be available for download from CCAFS at CIAT during March 2018. As they are a



highly specialised format, a Fortran module, MetGrid_Handler, has been produced for general distribution to assist their use. The full final report of the project can be found at Annex 5.

5.c *Outputs*

During March 2019, the compressed CLI files and the panes descriptor will be installed and made available from a world map interface at CIAT.

6. Better Insights into Germplasm through IWIN (BIGIWIN)

Urs Schulthess and Wei Xiong, CIMMYT, China.

6.a. *Project description*

The [International Wheat Improvement Network \(IWIN\)](#), as well as contributing to over half of the wheat germplasm grown worldwide, has amassed millions of yield and other agronomic data points from the yield and screening nurseries grown over approximately 4 decades by hundreds of public and private breeders globally. Relatively little of this data has been explored and used, due in part to a lack of metadata, such as daily temperature, radiation and other essential environmental information. The nursery trials are being conducted by many national programs worldwide. While there is a standard protocol in place, which encourages the partners to collect key parameters, not all of the sites follow the entire protocol. An initial analysis showed that key parameters, such as anthesis date are missing for some sites. Weather data are largely missing or not consistent. CIMMYT recently signed a data sharing agreement with meteoblue <www.meteoblue.com>, one of the leading providers of weather data at a global scale. It will provide CIMMYT with hourly weather data for the main nursery sites starting from 1986 until now. The data have a spatial resolution (grid size) that ranges between 8 and 30 km, depending on the



THE COMMUNITY OF PRACTICE ON **CROP MODELING**



geographic region. The main activity of this project will be to clean and curate the IWIN data sets and to fill gaps, such as anthesis date and other missing parameters, which will be modeled, to have a complete and rich dataset ready to be used for the crop modeling community and other research activities. This project will also use machine learning algorithms in order to establish a connection between phenotypes, genomic data and performance of CIMMYT's nursery lines at a global scale.

The outputs of the project will include:

- Quality controlled and gap filled IWIN nursery data sets starting from 1986 until 2018. Key phenology parameters, where missing, such as sowing, flowering and maturity date will be modeled and added to the dataset.
- IWIN data sets will be curated and made publically available via the GARDIAN portal following the CGIAR data sharing standards
- Hourly weather data for the key IWIN data sites (max 500) will be made publicly available, following the CGIAR data sharing standards. Duration: 1986 until 2018.

The project will start in February 2019 and will finish by January 2021.

6.b. Results

This project starts on February 2019, thus no results can be reported during 2018.

ANNEXES

Annex 1: Final report CGIAR Data Search for Crop Modeling

Name, position and affiliation of lead CG collaborators:

Senthod Asseng, University of Florida, USA
Gerrit Hoogenboom, University of Florida, USA
Cheryl Porter, University of Florida, USA
Chris Villalobos, University of Florida, USA

Name, position and affiliation of lead partner (if any):

Dr. Matthew Reynolds, CIMMYT
Dr. Diego Pequenno, CIMMYT

Start/end date: January to June 2018

Executive Summary

There is a wide variation in datasets available in GARDIAN, including the data quantity, quality, format, and terminologies used to describe the data. This project highlights three very different types of datasets and the steps involved to locate, access, and prepare the data for crop models using AgMIP data translation tools. Three datasets were retrieved from GARDIAN and evaluated for usefulness in crop modeling applications.

The current GARDIAN system is easy to navigate and to perform simple queries on data and publications; however there is a challenge to making these data useable on a large scale. Currently datasets are stored in many formats using vocabularies to describe dataset content which are determined on an ad hoc basis by each researcher. When the ICASA Data Dictionary is used as the definition of terms in the dataset, existing AgMIP data translation tools allow rapid translation of data to crop model-specific formats for multiple crop models. Making data useful on a large scale would require annotation of each dataset with terms and definitions that are in alignment with ICASA terms.

Data could be exposed to ICASA terminology by providing a mapping of variables in the dataset to terms in a standard ontology such as the AgrO. This ontology provides descriptions of each term and has been mapped to the ICASA Data Dictionary for selected terms. We recommend developing these annotations and variable mapping for datasets, and exposing the list of variables such that appropriateness for use in crop models can be easily determined with a simple query in GARDIAN. This proposed work was beyond the scope of this pilot project, but could be explored in a follow-up project.

Background

The CGIAR has a vast asset of field experimental data that could be extremely useful for crop model evaluation and improvement. In return, well-tested crop models can guide research in the CGIAR, particularly on adapting germplasm to future climate change and on developing crop management practices for sustainable and environmentally friendly food production. Under the “Organize” component of the Platform for Big Data in Agriculture, the CeRES tool (now renamed GARDIAN) is being developed to provide access to all the data that have been collected at the CGIAR Centers since their inception. The main goal of this proposal is to employ the current stage of this search engine to search, retrieve and translate the data into AgMIP format field experimental data sets. The specific objective is to develop an end-to-end case to show that the data are Findable, Accessible, Interoperable and Re-usable (FAIR). As a case study, we will explore the CIMMYT wheat data base for field data on germplasm with increased heat stress tolerance. To be useful for crop modeling, such field data need to include information on crop management (sowing, N fertilizer, irrigation, any effect from pest and diseases, lodging), soil (texture, soil organic carbon, bulk density, any constraints to root growth, potential rooting depth, and initial conditions), germplasm characteristics (phenology and yield characteristics, and specific traits, especially those related to heat tolerance), weather (daily max and min temperature, rainfall, solar radiation) and crop measurements (yield, yield components, phenology stages, growth patterns, soil water dynamics).

Outputs

- 1) An example of field data sets derived from the CGIAR data base (in this case for wheat on improved heat stress tolerance) in AgMIP format for model evaluation and improvement,
- 2) An evaluation of the currently developed data search engine for the Big Data Platform, and
- 3) An assessment of potential usefulness of CGIAR field data for crop modeling.
- 4) The results will be summarized in a report and scientific publication.

Deliverables

- 1) An example of field data sets derived from the CGIAR data base (in this case for wheat on improved heat stress tolerance) in AgMIP format for model development and testing.
- 2) An evaluation of the currently developed data search engine CeRES for the Big Data Platform,
- 3) New tools to link CeRES with the AgMIP data translators
- 4) An assessment of potential usefulness of CGIAR field data for crop modeling,
- 5) A summary report and scientific publication.

Deliverable 1 - Description and location of datasets

Three datasets were identified using the GARDIAN search engine for use as demonstration datasets for translation to crop modeling formats. The utility for crop modeling and the ease of translation to AgMIP and to crop modeling formats varied significantly between the three as discussed below. Additional details of each dataset are included in Appendix A.

- 1. International Heat Stress Genotype Experiment for modeling wheat response to heat: field experiments and AgMIP-Wheat multi-model simulations (GARDIAN ID: 543)**

This dataset was originally part of the AgMIP¹ Wheat Model Intercomparison. It includes basic modeling information such as experiment location, planting dates, soil profiles and initial conditions of the soil, and weather data. The data were stored in an Excel file which is formatted for the AgMIP translators using ICASA² variable annotations. The Excel spreadsheets, once converted to CSV files, are supported by existing AgMIP translators and therefore were easily converted to model formats. The weather files are also supported by an existing AgMIP translator. The data provided were sufficient to run a crop model with minimal use of AgMIP DOMEs³ to add model specific variables.

2. Calibrated DSSAT models for wheat (GARDIAN ID: 2020)

ICARDA⁴ provided this DSSAT⁵ wheat and cotton dataset for Uzbekistan. It contains DSSAT model specific information, such as cultivar parameters, as well as all the general modeling data; daily weather, soil profiles and initial conditions, and management information. The data were stored as DSSAT model files. By using the existing DSSAT translators, we are able to convert these to the model agnostic AgMIP formats and then to various model formats.

3. Wheat Derived Synthetics Heat Tolerance, Drought Tolerance, and Yield Potential Evaluation 2011-2014 (GARDIAN ID: 1497, 1538)

This dataset is a variety trial dataset which does not include sufficient data for modeling.

Deliverable 2 - Evaluation of GARDIAN search engine

The GARDIAN search engine works well for finding generic terms about a dataset, which have been previously added to the descriptions. It was simple to find the correct datasets specified for this project by keywords. However, GARDIAN currently is lacking support for more detailed searches when it concerns any data not contained in the tags, description or title of the dataset. For example, the search for *Wheat Uzbekistan* does not currently bring up the 2nd targeted dataset. This may be due to a lack of metadata for each dataset and the search therefore not leveraging that metadata, which will be addressed over time. Currently, this makes searching for quantitative and modeling datasets more difficult.

The output of the search is currently organized by the internal storage ID and unsortable (by date for instance), however the filters work well for narrowing down some of the searches. An individual dataset, once selected by the search results, can be difficult to navigate. The **Dataset Files**

¹ Agricultural Model Intercomparison and Improvement Project (<http://www.agmip.org/> Rosenzweig et al., 2012).

² International Consortium for Agricultural Systems Applications, (White et al., 2013).

³ AgMIP DOME files supply information required by crop models, but not included in the original dataset.

⁴ International Center for Agricultural Research in the Dry Areas

⁵ Decision Support System for Agrotechnology Transfer

dropdown properly shows a list of downloadable files from the dataset, but the link to the files themselves are not apparent from there. Also, the various links themselves can be confusing for new users. The HDL, URL, DOI terms could be expanded to a *Download Data (DOI)* for example.

Deliverable 3 - Flexible format data translation tool

As part of the AgMIP data interoperability system (Porter et al., 2014), a few generic templates for data entry were developed for commonly encountered data sources for modeling. These include farm survey data, detailed field crop research data, paddy rice systems, and others. Each template is an Excel spreadsheet with representative data fields annotated with the ICASA variable name for each and have encoding to help the translate determine how data are relationally connected. This approach requires that spreadsheets be saved in comma-delimited format (csv) translated using java desktop utilities.

A new application was developed that allows Excel spreadsheets to be read directly and relational connections be automatically discovered by the application. This allows the data manager much greater flexibility in organization of the data and also removes the burden of creating csv files from the Excel workbook prior to translation.

Each data element in the dataset must still be mapped to an ICASA variable in order for the translator to interpret the values and produce meaningful crop model inputs. This can be a time consuming task and must be done by someone familiar with the data set, the methods of data collection, and the ICASA Data Dictionary.

Deliverable 4 - Potential usefulness of GARDIAN for crop modeling

We found that data in GARDIAN were well annotated and easily found with a few simple keywords. For example, using the keywords “wheat” and “heat” brought up 8 datasets as shown in Figure 1 below. Selection of the 2nd dataset, “Data from the International Heat Stress Genotype Experiment for modeling wheat response to heat: field experiments and AgMIP-Wheat multi-model simulations”, brought up a summary of the data, a word cloud of keywords, a list of relevant publications, and a link to the dataset in compressed “rar” format. The link leads to the Open Data Journal for Agricultural Research (ODJAR , <https://library.wur.nl/ojs/index.php/odjar/>) from which additional descriptions of the dataset and the data are available.

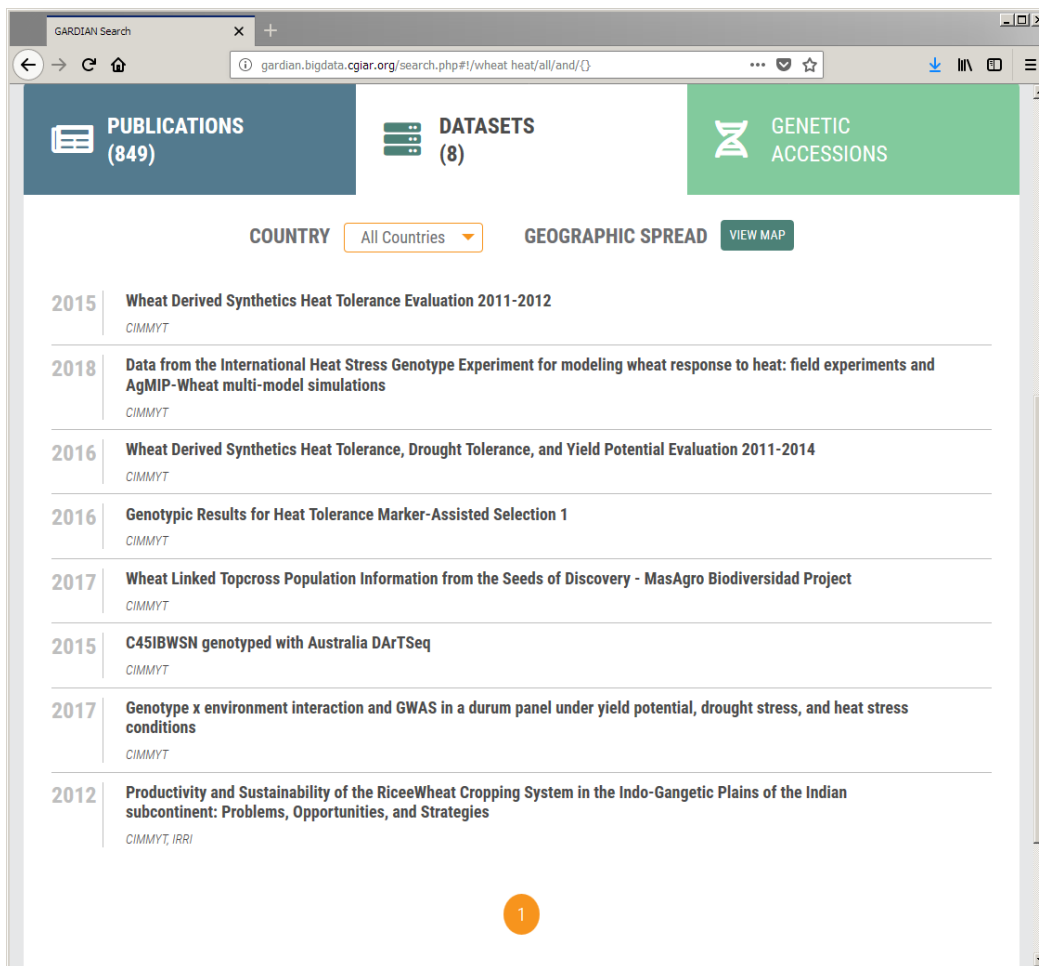


Figure 1. Screen shot of results of search in GARDIAN for “wheat” and “heat”.

While GARDIAN is a very user-friendly method of discovering data that are available, there is a challenge to making these data useable on a large scale. Currently datasets are stored in many formats using vocabularies to describe dataset content which are determined on an ad hoc basis by each researcher. The Big Data Platform proposes to align each dataset vocabulary to a well annotated vocabulary that is described in the Agronomy Ontology (Devare, et al., 2016). Vocabulary definitions for each dataset are to be described fully in metadata so that researchers can maintain their own vocabularies and methods of data collection and description, while at the same time, linking these data to other datasets through mapping of variables to a well-documented and maintained ontology. Once data are annotated with these metadata, the datasets are made not only findable and accessible, but interoperable. Combining the annotated datasets with AgMIP data translation tools, makes these data further usable by allowing them to be used in crop modeling applications associated with AgMIP.

Deliverable 5 - summary report and scientific publication

This document consists of the final summary report for the project. A scientific publication will be written as part of a follow-up project with the CGIAR which integrates AgMIP data translation tools in the GARDIAN user interface.

Visibility of output (in terms of reinforcing the value of the CoP and Big Data Platform)

The field data sets have been made available to the international crop modeling community through upload to the AgMIP Crop Site Database. AgMIP is one of the largest international research communities with more than 1000 members. The Big Data Platform will be highlighted through this wheat example. This will ensure great visibility of the potential impact of the Big Data Platform in CGIAR for other scientific communities, with broader implications for future research direction and policy makers.

The impact of the proposed project is awareness to the crop modeling community of the CGIAR Big Data Platform and the FAIR principles of CGIAR data, and the broader support and impact of the CGIAR to science and policy.

References

- Devare, M., Aubert, C., Laporte, M.-A., Valette, L., Arnaud, E., Buttigieg, P.L., 2016. Data-driven agricultural research for development - a need for data harmonization via semantics. In: Jaiswal, P., Hoehndorf, R. (Eds.), 7th International Conference on Biomedical Ontologies, ICBO'16, vol. 1747 of CEUR Workshop Proceedings, Corvallis, Oregon, USA, pp. 2, August.
- Porter, C.H., C. Villalobos, D. Holzworth, R. Nelson, J.W. White, I.N. Athanasiadis, S. Janssen, D. Ripoche, J. Cufi, D. Raes, M. Zhang, R. Knapen, R. Sahajpal, K.J. Boote, J.W. Jones. 2014. Harmonization and translation of crop modeling data to ensure interoperability. Environmental Modelling and Software. 62:495-508. [doi:10.1016/j.envsoft.2014.09.004](https://doi.org/10.1016/j.envsoft.2014.09.004).
- Rosenzweig, C., J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P. Thorburn, J.M. Antle, G.C. Nelson, C. Porter, S. Janssen, S. Asseng, B. Basso, F. Ewert, D. Wallach, G. Baigorria, and J.M. Winter. 2012. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and Pilot Studies. Ag. For. Meteorol. 170:166-182 <http://dx.doi.org/10.1016/j.agrformet.2012.09.011>.
- White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W., Hoogenboom, G. 2013. Integrated Description of Agricultural Field Experiments and Production: the ICASA Version 2.0 Data Standards. Computers and Electronics in Agriculture. 96:1-12.

Appendix A. Conversion of data to modeling formats

Dataset 1. International Heat Stress Genotype Experiment for modeling wheat response to heat: field experiments and AgMIP-Wheat multi-model simulations

This dataset includes 28 treatments from six locations that were used in an AgMIP wheat intercomparison exercise. The data were archived on the Open Data Journal for Agricultural Research (<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CJBSR>). These data were archived in AgMIP format, although the file was stored as an xml file, rather than an xlsx file. After converting from xml to xlsx, thence to csv files, the AgMIP data translations tools could be used to translate to DSSAT model format. Figure A-1 illustrates graphical outputs from the DSSAT model showing aboveground biomass and grain weight in kg/ha for 4 of the 28 treatments included in the dataset. Table A-1 presents a list of variables included in this dataset.

This exercise revealed a potential error in the archived data. A single soil profile was archived for Maricopa, Arizona, USA, rather than site specific soils for the six field locations. The soil profile data archived may be a remnant of a data template used to input data to AgMIP format. Therefore the results of the crop model simulations are not considered to be reliable and were not compared to observed data.

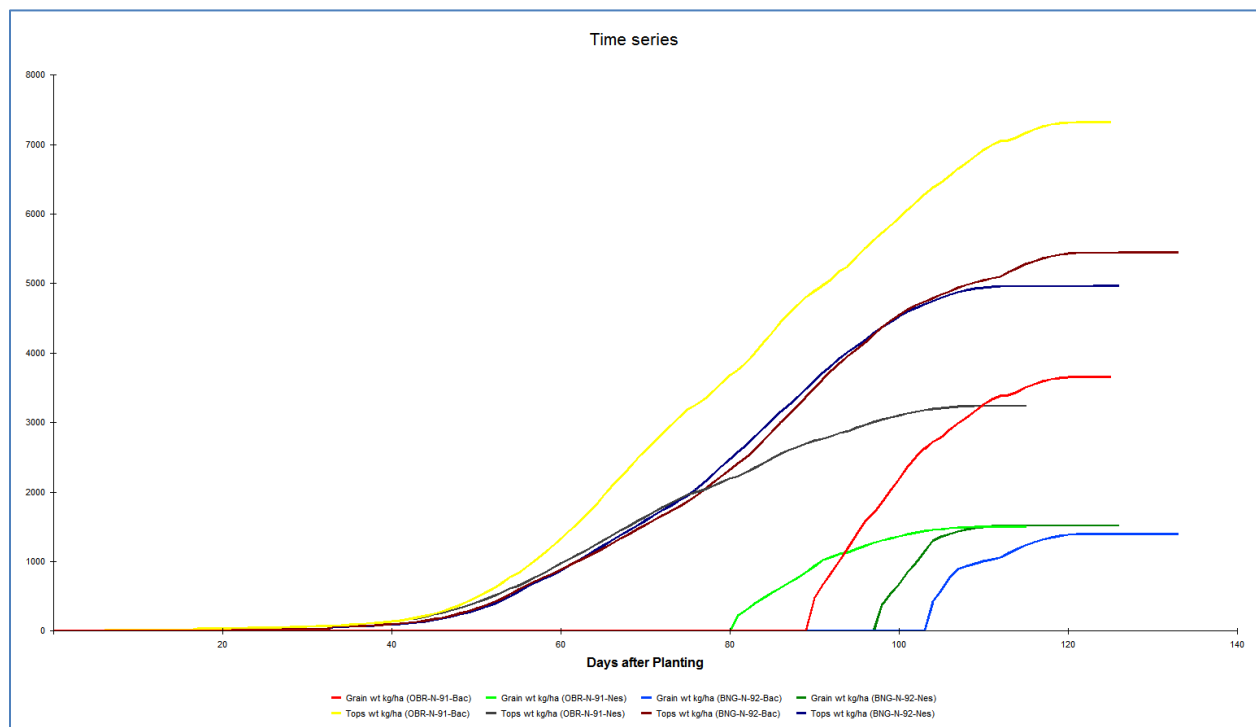


Figure A-1. Simulated outputs from International Heat Stress Genotype Experiment for modeling wheat response to heat dataset showing aboveground biomass (Tops wt, kg/ha) and grain weight (Grain wt, kg/ha) for four treatments.

Table A-1. Variables included in from International Heat Stress Genotype Experiment for modeling wheat response to heat dataset

Variable	Definition	Units
EXNAME	Experiment name	
TRT_NAME	treatment name	
LOCAL_NAME	Location	
EXP_DUR	Duration of experiment	year
FL_LAT	Latitude	decimal degrees
FL_LONG	Longitude	decimal degrees
FLELE	Field Elevation	m
ICDAT	Initial conditions date	yyyy-mm-dd
ICPCR	Previous crop code	Wheat (WH)
ICRAG	Initial surface residue	g[dry matter]/m2
ICRT	Initial root residue	g[dry matter]/m2
ICRN	Initial residue N conc	%
ICBL	Soil layer base depth	cm
ICH2O	Initial water content	mm3/mm3
ICNH4	Initial ammonium conc	ppm
ICNO3	Initial nitrate conc	ppm
PDATE	Sowing date	date
CRID	Crop ID	code
CUL_ID	Cultivar ID	code
CV_NOTES	Additional data about cultivar	text
PLPOP	Plant Population	#/m2
PLRS	Row spacing	cm
PLDP	Planting depth	cm
SLLB	Soil layer base depth	cm
SLLL	Lower limit	mm3/mm3
SLDUL	Drained upper limit	mm3/mm3
SLSAT	Saturation	mm3/mm3
SKSAT	saturated hydraulic conductivity	cm/h
SLBDM	Bulk density	g/cm3
SLOC	Organic C	g[C]/100g
SLCL	Soil texture, clay (<0.002 mm)	%
SLSI	Soil texture, silt (0.05 to 0.002 mm)	%
SLCF	Soil texture, coarse fraction (>2 mm)	%
SLCEC	Cation exchange capacity	cmol/kg
SRAD	Solar radiation	Temperature of air,

		maximum
TMAX	Maximum daily temperature	°C
TMIN	Minimum daily temperature	°C
RAIN	Daily precipitation	mm
WIND	Daily wind run	km/d
TDEW	Temperature, dewpoint	°C
VPRSD	Vapor pressure, average daily	kPa
RHUMD	Relative humidity at maximum daily temperature (or minimum daily relative humidity)	%
HDATE	Harvest date	
PDATE	Planting date	
PLDAE	Emergence date	
ADAT	Anthesis date	
MDAT	Maturity date	
Tea *	Average daily mean air temperature from crop emergence to anthesis	°C
Taegf *	Average daily mean air temperature from anthesis to maturity	°C
Teegf *	Average daily mean air temperature from crop emergence to maturity	°C
HWAM	Yield	t/ DM/ha
hwams *	Yield.sd	t/ DM/ha
CWAA	Biom.an	t/ DM/ha
cwaas *	Biom.an.sd	t/ DM/ha
CWAM	Biom.ma	t/ DM/ha
CWAMS	Biom.ma.sd	t/ DM/ha
HIAM	HI	-
hiams *	HI.sd	-
H#AM	GNumber	grain/m ²
h#ams *	GNumber.sd	grain/m ²
GWGM	GDM.ma	mg DM/grain
gwgms *	GDM.ma.sd	mg DM/grain
E#AD	NEM2	ear/m ²
e#ads *	NEM2.sd	ear/m ²
G#SD	NGE	grain/ear
g#sds *	NGE.sd	grain/ear
PLPAM	NPM2	plant/m ²
plpams *	NPM2.sd	plant/m ²
EHTX	Plant.height	m
ehtxs *	Plant.height.sd	m

* Variable is not in ICASA Data Dictionary

Dataset 2. ICARDA Calibrated DSSAT models for wheat

This dataset consists of cotton and wheat data archived on the CGIAR Monitoring, Evaluation, & Learning Repository (<http://repo.mel.cgiar.org/handle/20.500.11766/3537>). Two wheat irrigation treatments and three N fertilizer cotton treatments are included from experiments in Uzbekistan.

Because this dataset was archived in DSSAT format, very little time was required to prepare the files for simulation for both wheat and cotton. Model-specific species, ecotype and cultivar files had to be updated for the latest version of DSSAT-CSM and the soil file had to be renamed to conform to DSSAT file naming conventions. Although the cultivar was calibrated for DSSAT v4.6, the results are acceptable when run with DSSAT v4.7 for the wheat data (Fig A-2). The cotton data (Fig A-3) do not compare as favorably to observed data, and possibly require additional calibration for DSSAT-CSM v4.7.2.

Table A-2 presents the variables included in the dataset.

Although the archived files were formatted for DSSAT, these files have been converted to AgMIP format and are available for translation to other model formats.

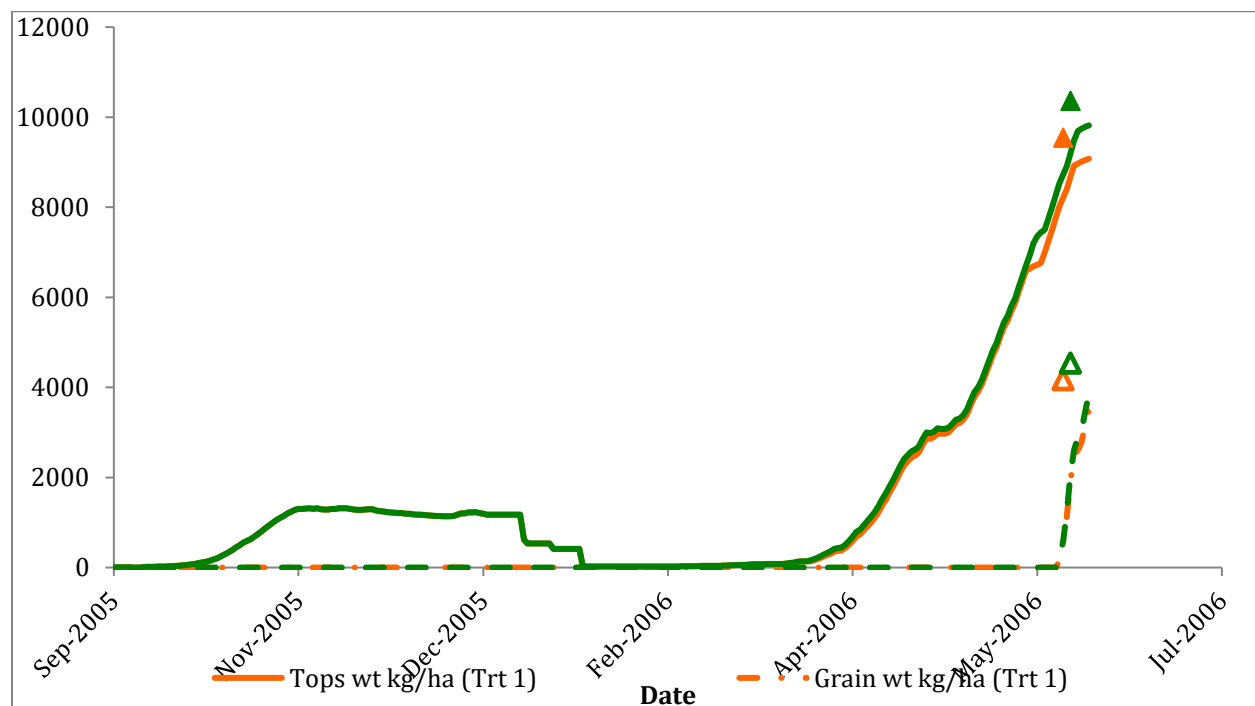


Figure A-2. Simulated and observed biomass (tops weight) and grain weight for two wheat treatments included with data set 2, “ICARDA Calibrated DSSAT models for wheat”.

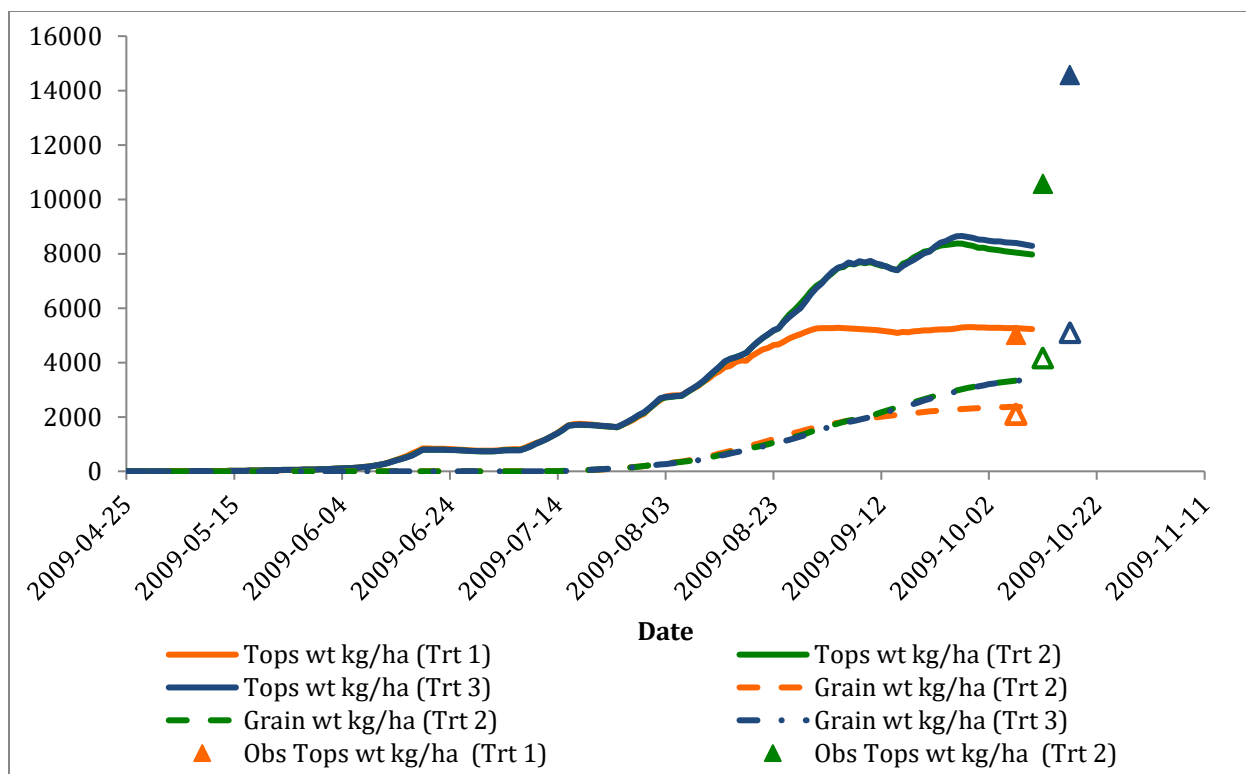


Figure A-3. Simulated and observed biomass (tops weight) and grain weight for two wheat treatments included with data set 2, “ICARDA Calibrated DSSAT models for wheat”.

Table A-2.

Variable	Definition	Units
CRID	Crop (or weed) species identifier	
CUL_ID	Cultivar, line or genotype identifier	
CUL_NAME	Cultivar name	
PDATE	planting_date	date
EDATE	Date of emergence	date
PLDP	planting_depth	mm
PLRS	row_spacing	cm
PPOP	plant_pop_at_planting	number/m2
PLPOE	Plant population at emergence	number/m2
PLMA	Planting material	
PLDS	Planting distribution	
SH2O	initial_watr_conc_by_lyr	cm3/cm3
SNH4	initial_NH4_concen_layer	ppm
SNO3	initial_NO3_concen_layer	ppm
IDATE	Irrigation date	date
IREFF	Irrigation application efficiency as fraction (0 to 1)	

IROP	Irrigation operation (e.g., furrow, sprinkler, drip...)	
IRVAL	Irrigation amount, depth of water	mm
FEDATE	Fertilizer application date	date
FECD	Fertilizer material	
FEACD	Fertilizer application method	
FEDEP	Fertilizer application depth	cm
FEAMN	Nitrogen in applied fertilizer	kg[N]/ha
FEAMP	Phosphorus in applied fertilizer	kg[P]/ha
FEAMK	Potassium in applied fertilizer	kg[K]/ha
TDATE	Tillage date	date
TIIMP	Tillage implement	code
TIDEP	Tillage operations depth	cm
SALB	soil_albedo	fraction
SLDR	Drainage rate as fraction per day	1/d
SLRO	Runoff curve no. (Soil Conservation Service)	number
SMPX	Phosphorus determination method	
SLLB	Soil layer base depth	cm
SLBDM	soil_bulk_density_moist	g/cm3
SLCL	soil_clay_fraction	%
SLSI	soil_silt_fraction	%
SLLL	soil_water_lower_limit	cm3/cm3
SDUL	soil_wat_drned_upper_lim	cm3/cm3
SSAT	soil_water_saturated	cm3/cm3
SSKS	sat_hydraul_conductivity	cm/h
SLRGF	Root growth factor, soil only (0 to 1 scale)	number
SLOC	soil_organic_C_perc_layr	g[C]/100g[soil]
SLNI	Nitrogen, total soil organic	%
SLPHW	Soil pH in water	
SLPX	Phosphorus, extractable	mg/kg
w_date	weather_date	date
wst_lat	weather_sta_latitude	decimal degrees
wst_long	weather_sta_longitude	decimal degrees
SRAD	Solar radiation	Temperature of air, maximum
TMAX	Maximum daily temperature	°C
TMIN	Minimum daily temperature	°C
RAIN	Daily precipitation	mm
WIND	Daily wind run	km/d

RHUMD	Relative humidity at maximum daily temperature (or minimum daily relative humidity)	%
EDAT	Emergence date	date
ADAT	Anthesis date	date
MDAT	Maturity date	date
HWAM	Harvest weight at maturity	kg/ha
CWAM	Biomass at maturity	kg/ha
BWAH	Harvested byproduct weight	kg/ha
HIAM	Harvest index	
LAIX	Maximum leaf area index	m ² /m ²
HWUM	Harvest unit dry (e.g., seed) weight at harv. maturity	g/unit
H#AM	Harvest number per area at maturity (e.g., seed or tubers)	number/m ²
CHTA	Canopy height at anthesis	m
GN%M	Grain N concentration at maturity	%
GNAM	Grain N at maturity	kg/ha
SNAM	Stem N at maturity	kg/ha
CNAM	Nitrogen in above ground plant parts at maturity	kg/ha

Dataset 3. Wheat Derived Synthetics Heat Tolerance, Drought Tolerance, and Yield Potential Evaluation 2011-2014

This dataset was archived on the CIMMYT Maize-Wheat Data Repository at <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10394> and <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10656>.

Table A-3 lists the variables included in the dataset. There were insufficient data to be relevant to crop models. A minimum set of crop model-relevant data would include management operations data (including planting date, row spacing, planting density, amounts and dates of fertilizer applications, amounts and dates of irrigation application, etc.); soil property data; and observations of plant growth and development (including grain weight over time, biomass weight over time, anthesis date, maturity date, leaf area index, etc.).

Table A-3. Variable listing for Dataset 3

Column Header	Data type	Description	Additional information
Ent.	Text	Entry number	
Row	Integer	Row in field	Y coordinate
Column	Integer	Column in field	X coordinate
CID	Integer	Cross ID	CIMMYT IWIS2 ID
SID	Integer	Selection ID	CIMMYT IWIS2 ID
GID	Integer	Germplasm ID	
Cross Name	Text	Cross Name	

BW/Synth/Check	Text	Type of material	B=Bread wheat; S=Derived Synthetic; C=check
BP/BC/TC	Text	Type of cross Selection	BP=Bi-parental cross; BC=Backcross; TC=Top cross
Sel/Unsel	Text	decision	
BW Parent	Text	Bread wheat parent	
Synth Parent	Text	Synthetic hexaploid wheat parent Selection	
Sel_Hist	Text	History	
Origin	Text	Seed source	
Height (cm)	Integer	Plant height in centimeters	
Yield (T/ha)	Decimal	Grain yield in tons per hectare	
Comments	Text	Comments	Several abbreviations are used including: gd=good; yld=yield; v=very; hds=heads; b or br=brown
Germination Failure	Text	Observations concerning germination	
GID Y10-11		These GIDs link lines to marker file (DArT-seq GBS marker).	
Average_temperature (C)			
Maximum_temperature (C)			
Minimum_temperature (C)			
Precipitation (mm)			
Evaporation (mm)			

Annex 2

Development of system modeling platform to guide agro-ecosystem specific interventions to enhance post-rainy sorghum production in India

Partial Report 2018

Kholova, Jana (CIAT); Ronanki, Swarna (Indian Institute of Millet Research)

Project Status:

I. Insilico – experiment: Simulation of post rainy sorghum yield response to variable Nitrogen levels

Sorghum module of APSIM platform (v7.6) was used to simulate the yields of rabi sorghum in the districts of Karnataka, Maharashtra, Andhra Pradesh and Telangana; for each of the simulation units, the soil information was compiled from NBSS-LUP and ISRIC, 50 years of meteorological data was generated by the Marksim (v1.0), sorghum crop management reconstructed from the recommendations described in Trivedi et al. 2008 and Maldandi plant type coefficients transcribed from Ravi Kumar *et al.* 2009. The relevance of such set-up was tested against the observed yields and meteorological records before. (Kholova *et al* 2013, 2014; Swarna *et al.*, 2017).

The first set of baseline simulations was carried out with the recommended farmer practice (FP) of applying 20 kg urea/ha as a starter dose and 20 kg urea/ha as top dressing. The second set was carried out using the typical on-station N practice of applying 50 kg DAP/ha as a starter dose and 100 kg urea/ha as top dressing. Grain and stover yields generated under on-station N management were compared to the farmer practice and the frequencies of yield advantage of OS over FP were calculated for all simulation units and projected onto a map (Fig 1). The on-station practice enhanced stover production across central India, but in low yielding environments average yield was 1500 kg/ha with significant grain yield loss. This was on expected lines. Well-fertilized crops establish larger canopy earlier in the season which reflects in higher crop demand for transpiration and leads to earlier water depletion from the soil. Therefore, on-station crops face water stress earlier in the season with less moisture available to facilitate grain-filling processes.

Our study showed that out of 83 investigated districts, only 19 would benefit from OS-N practice in terms of sorghum production. These 19 districts generally encompass regions with deep vertisols with high soil water holding capacity; higher chance of in-crop rainfall – i.e. less chance of severe droughts. In other words, majority of post-rainy sorghum production areas will suffer the grain production losses if higher doses of N fertilizers were applied.

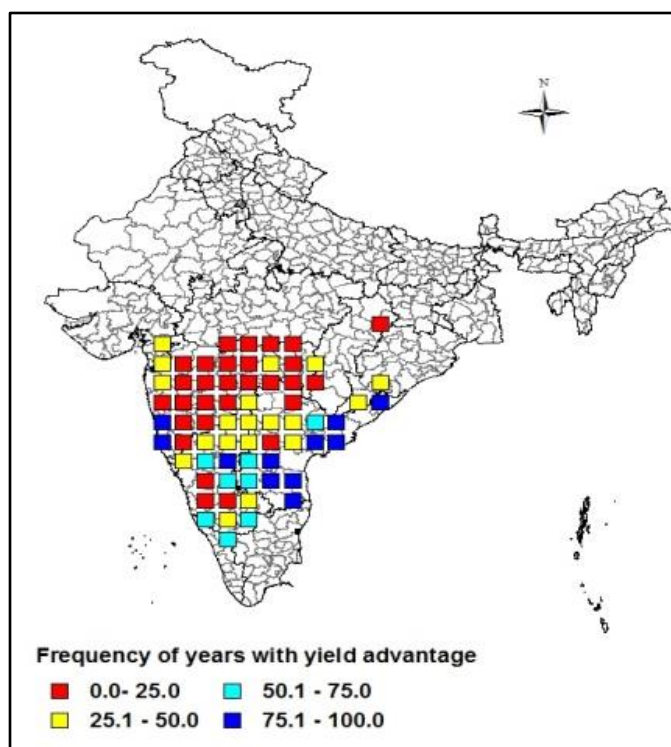


Fig 1. Frequency of years with grain yield advantage due to on-station N practice. The yellow and red squares signify higher likelihood of grain yield loss while the light and dark blue squares signify higher likelihood of grain yield gain due to on-station N application (OS).

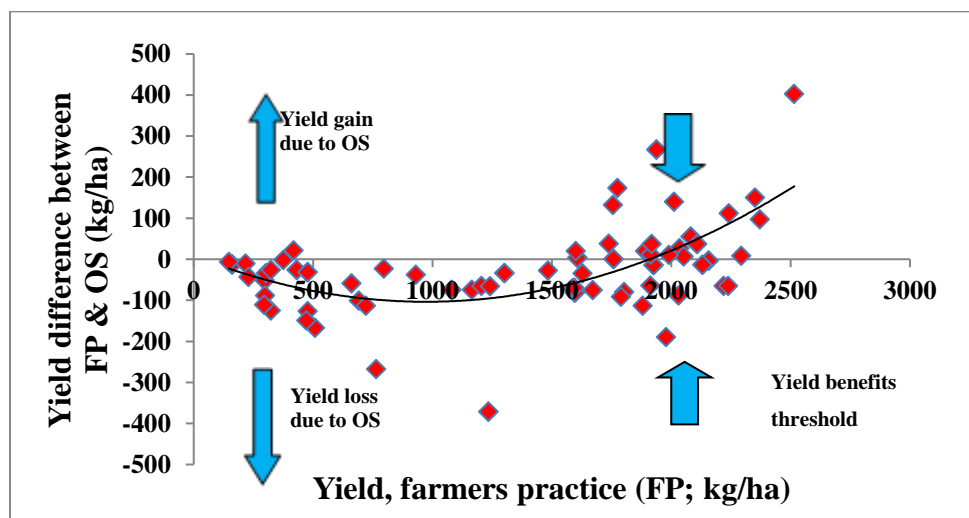


Fig 2: Yield difference between farmers practice (FP) and optimum N practice (OS) was negative in low yielding environments

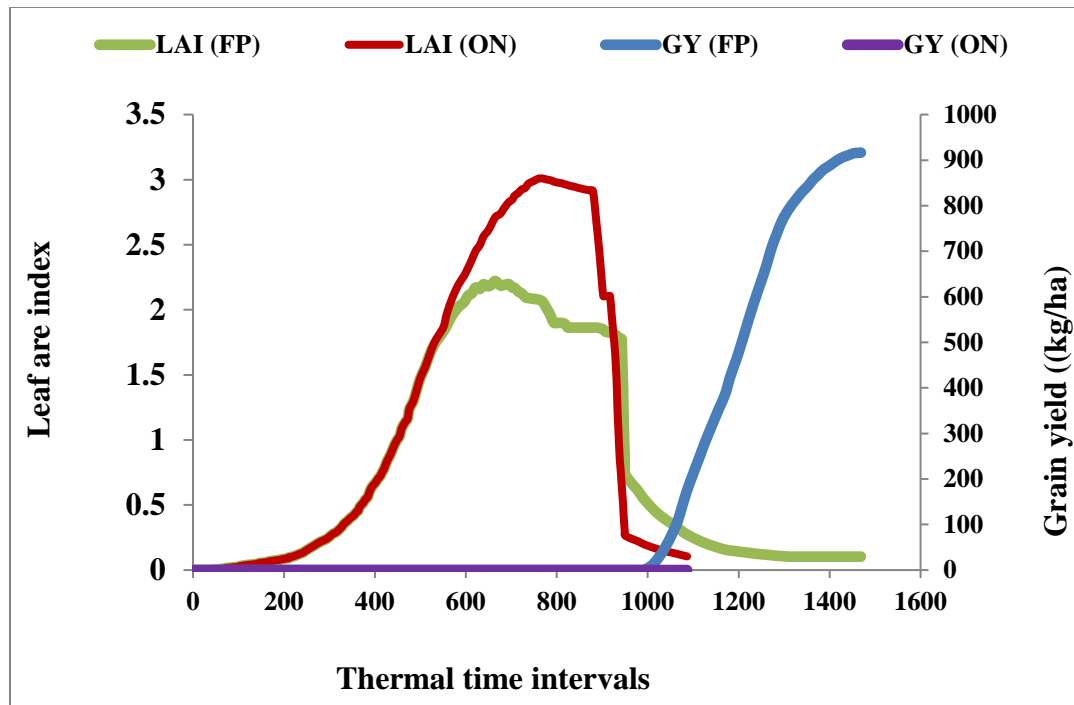


Fig 3: Simulation of canopy growth dynamics with low N application (orange line is farmer practice) and non-limiting N application (blue line is on-station N practice) along with associated grain yield (GY). The well-fertilized crop failed to yield (purple line) while the lower dose of N application resulted in crop yields of ~900 kg/ha (red line).

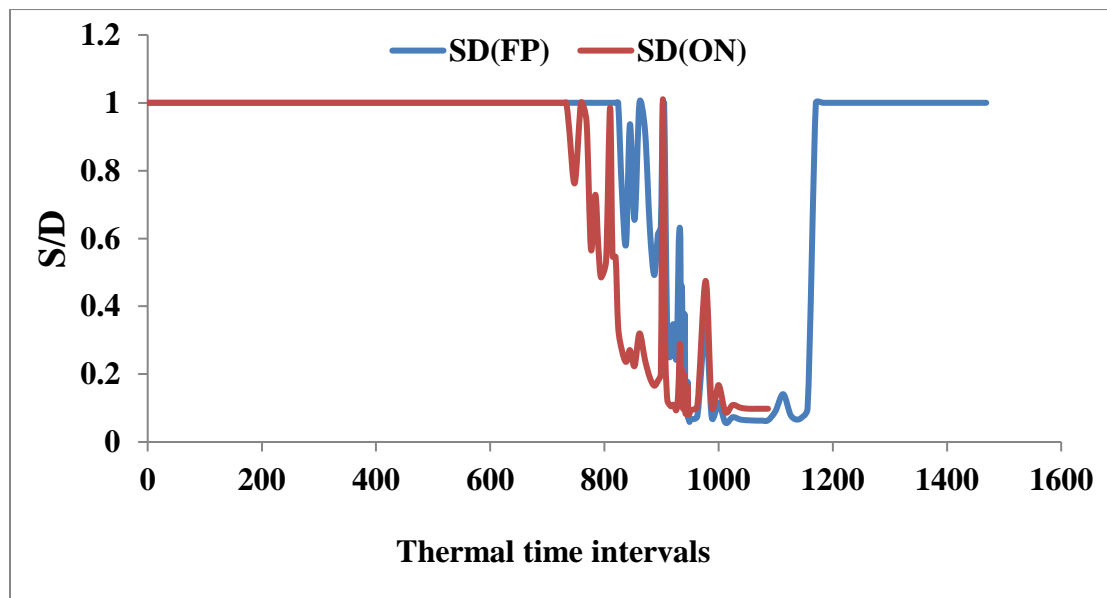


Fig 4: Simulation of crop water depletion dynamics during the season. S/D is the water supply/demand ratio (the lower the ratio, the larger the stress effect on the crop); the orange line stands for crop grown with limited N input (farmers practice) where stress begins later in the season compared to the blue line which is the water stress trajectory of the crop raised using on-station N practice.

II. Insilico – experiment: Simulation of post rainy sorghum yield response to variable plant densities

The simulations baseline was carried out considering the recommended planting density (RP; 12 plants per m^2) and second set of simulations were carried out with low density planting (LDP; 6 plants per m^2) and third set of simulations were carried out with high density planting (HDP; 18 plants per m^2). Grain and stover yields generated under these three simulations were compared and the yield difference and frequencies of yield advantage were calculated for all simulation units and projected onto a map using ArcGIS software (v. 10.3.1). The results indicated that there was an increased grain yield with adoption of 6 plants per m^2 in most of the districts over 12 plants per m^2 (Fig5a). Adoption of high density planting with 18 plants per m^2 resulted in decreased yields and even complete crop failure in most of the rabi sorghum areas (Fig 5b). The beneficial effect of increased plant density was realized in good environmental and soil conditions. Contrary to grain yield, the stover yield increased with increasing plant density from 12 to 18 plants per m^2 and decreased with reducing the plant density from 12 to 6 plants per m^2 .

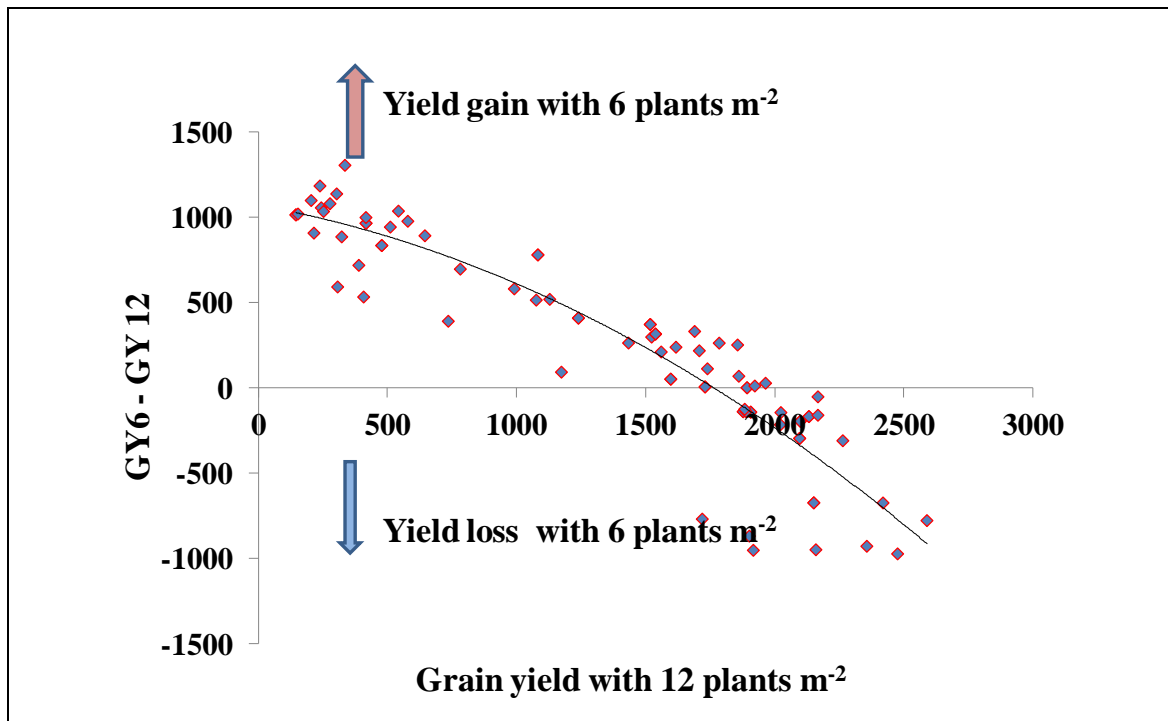


Fig. 5a. Grain yield difference: 6 plants m^{-2} versus 12 plants m^{-2}

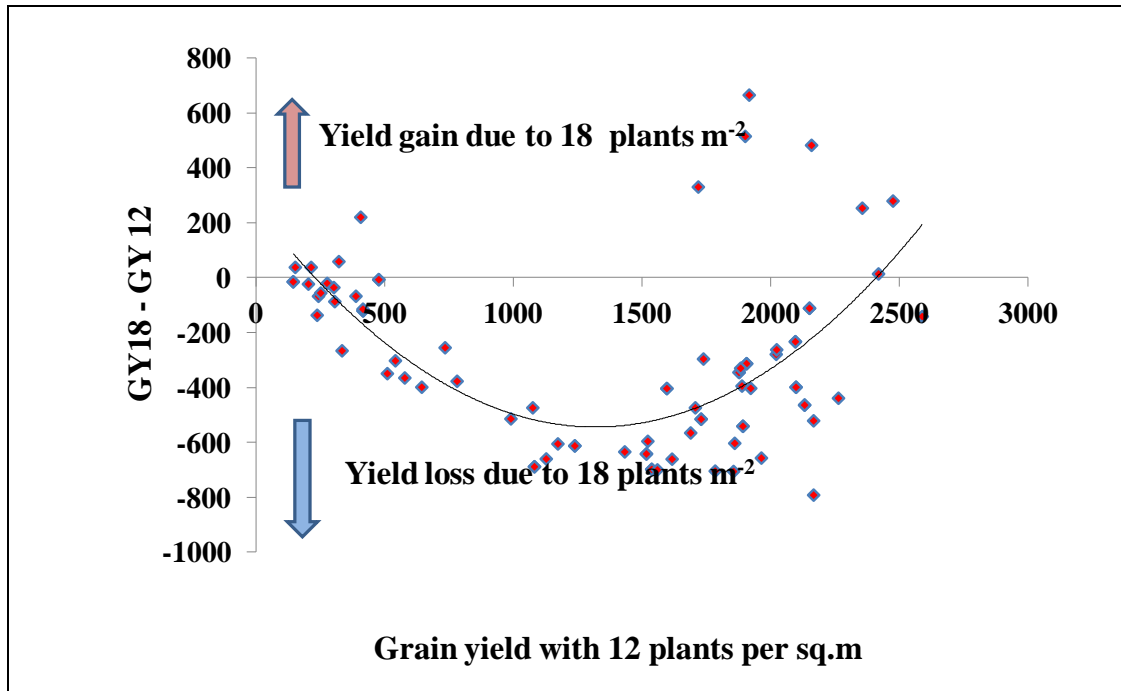


Fig. 5b. Grain yield difference: 12 plants m⁻² versus 18 plants m⁻²

III. Validation of weather data

In the insilico experiment, Marksim generated weather data was used with a resolution of 100 km. In order to increase the resolution of the existing APSIM simulation grid, NASA weather data which is of 50 km resolution (Fig 6a), was collected and validated against observed weather data. There was a quite good agreement between NASA and observed weather data with respect to maximum and minimum temperature (Fig 6b) and rainfall (Fig 6c)

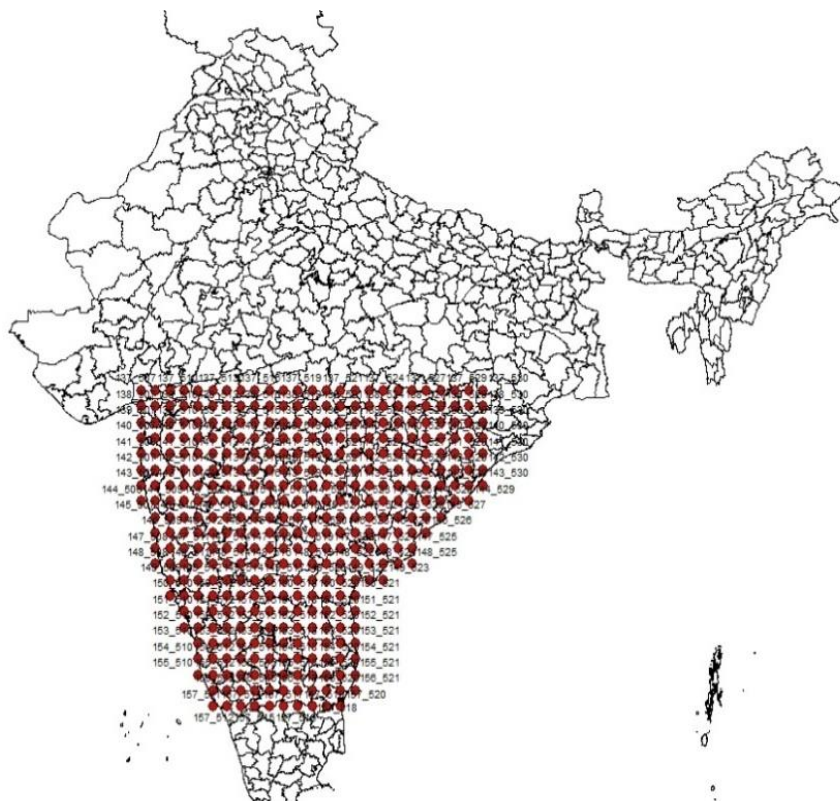


Fig 6a. NASA weather data (50 × 50 km)

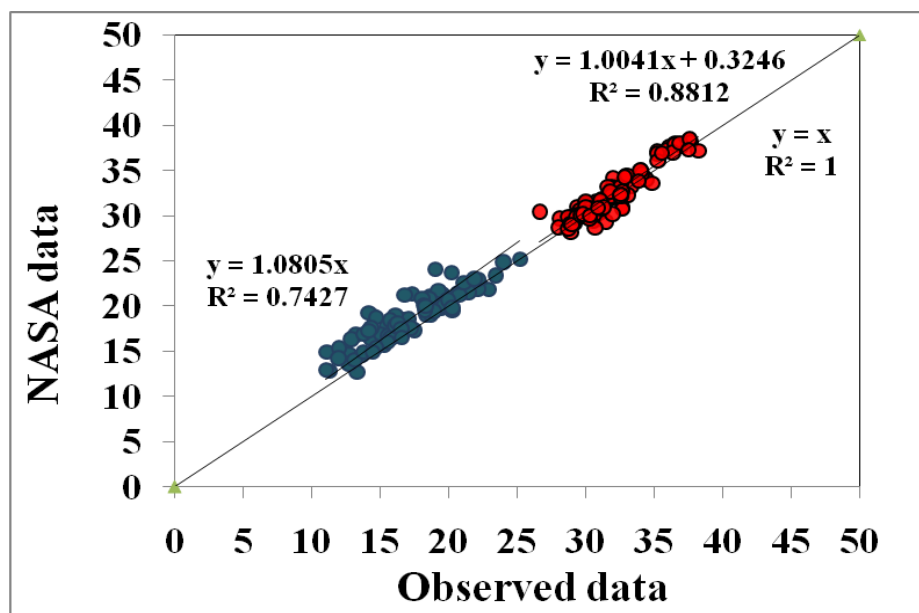


Fig 6b. Validation of NASA data with observed weather data for maximum and minimum temperature (°C)

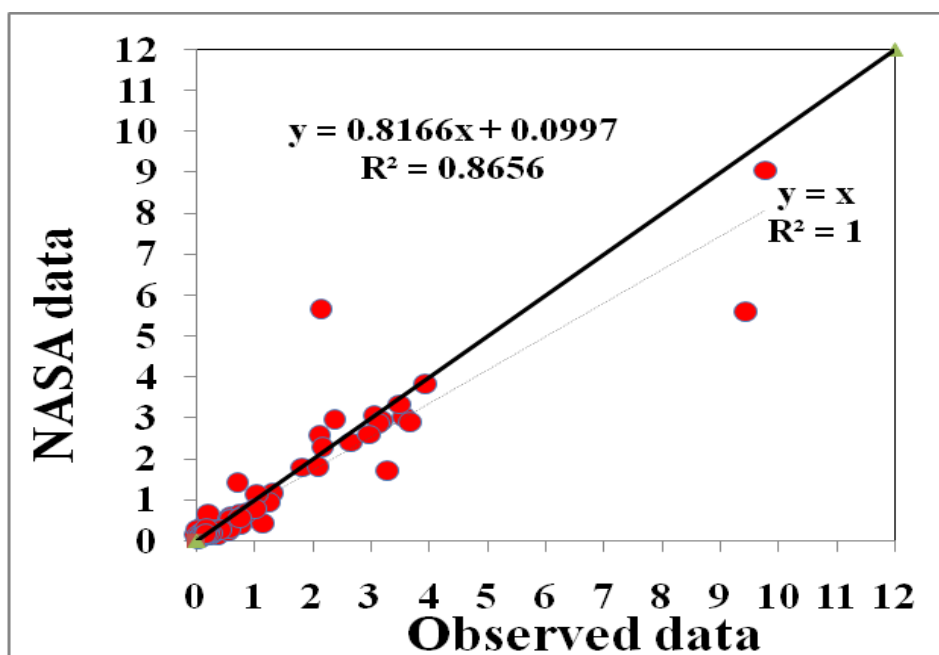


Fig 6c. Validation of NASA data with observed weather data for rainfall (mm)

Calibration and validation of APSIM:

- ✓ Field experiments were conducted in ICRISAT and ICAR- IIMR during rabi 2017-18 for collecting the data for parameterization of rabi sorghum cultivars into APSIM and for validation of APSIM

Table 1: List of rabi sorghum cultivars parameterized

✓ CSV 14R	✓ Phule Chitra
✓ CSV 18	✓ Phule Anuradha
✓ CSV 22R	✓ Phule Suchitra
✓ CSV 26R	✓ Parbhani Moti
✓ CSV 29 R	✓ Phule Vasudha
✓ CSV 216 R	✓ PKV Kranti
✓ M 35-1	✓ Phule Maulee
✓ Phule Revati	

- ✓ Experiments were conducted in split plot design in well watered and water stress conditions at two plant densities and two nitrogen levels

- ✓ Data on agronomic parameters (Phenology, biomass accumulation and grain yield) and canopy characteristics were documented to derive genetic coefficients.
- ✓ The observations necessary to derive reliable genetic coefficients for parameterization of cultivars was collected from low density, well-fertilized and irrigated treatment and will be validated upon other treatments.
- ✓ The detailed parameterization process, basic equations and coefficient optimization tools used are available open-access at:

<http://gems.icrisat.org/allinstruments/parameterization-for-apsim-simulation/>

Parameterisation of cultivars for canopy development

For parameterisation, the data was collected from irrigated, low density and high nitrogen treatment. Firstly, maximum leaf area for all the cultivars was determined by calculating the relation between maximum leaf number and maximum leaf size (Fig. 7)

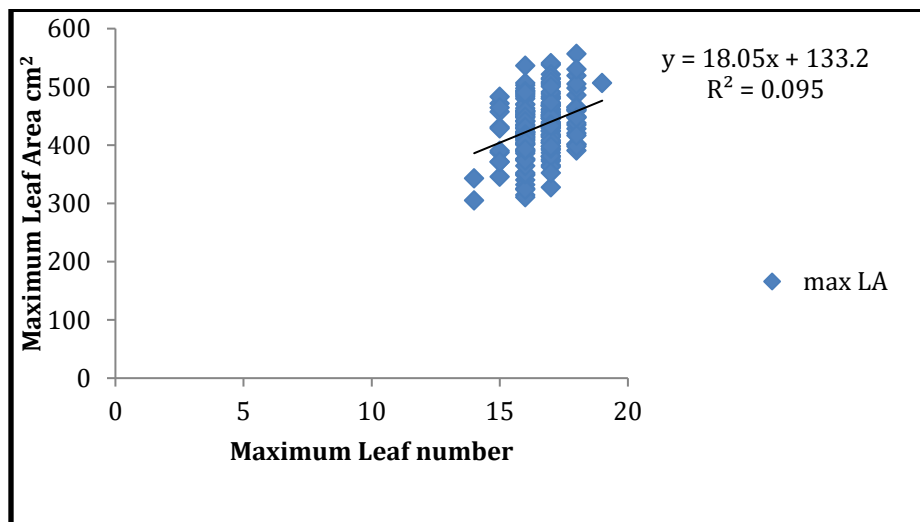


Fig 7. Relation between maximum leaf number and maximum leaf area for all the rabi sorghum cultivars.

For the ease of calculations, the cultivars are grouped into 4 types based on maximum leaf area values and canopy coefficients are derived for each group:

Group 1: Maximum leaf area - $< 400 \text{ cm}^2$
 Cultivars: Phule Anuradha, Phule Maulee

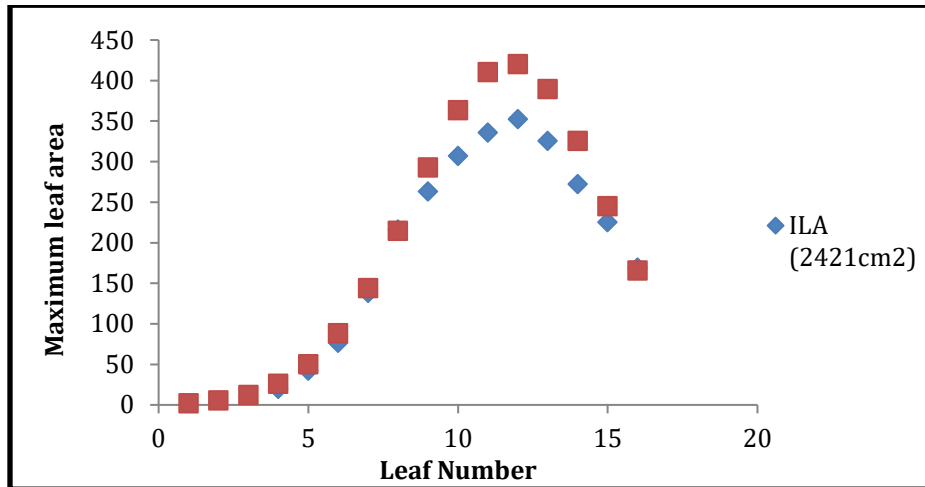


Fig 8a. Optimization for ILA function for cultivars Phule Anuradha and Phule Maulee

Group 2: Maximum leaf area – $400 - 420 \text{ cm}^2$

Cultivars: CSV 26, CSV 14R, Phule Vasudha, Phule Revathi

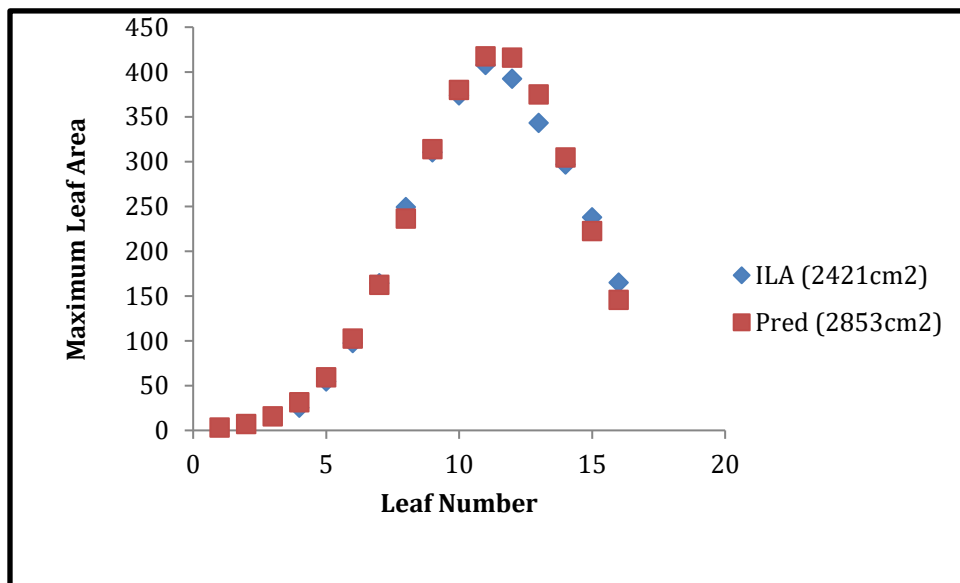


Fig 8b. Optimization for ILA function for cultivars CSV 26, CSV 14R, Phule Vasudha, Phule Revathi

Group 3: Maximum leaf area – 420 – 450 cm²

Cultivars: CSV 18, CSV 22, CSV 29 R, AKSV 13 R, M 35-1. CSV 216 R

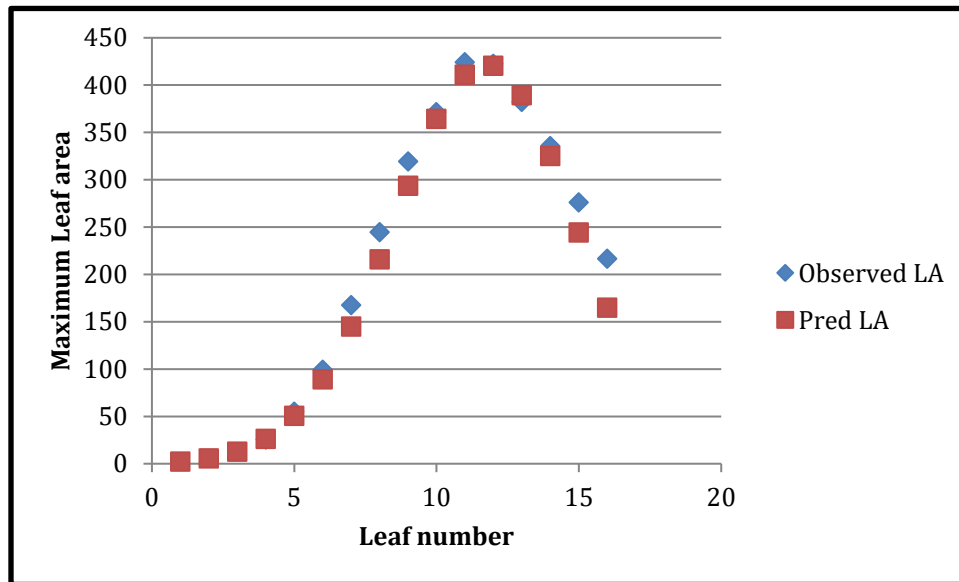


Fig 8c. Optimization for ILA function for cultivars CSV 18, CSV 22, CSV 29 R, AKSV 13 R, M 35-1. CSV 216 R

Group 4: Maximum leaf area – > 450 cm²

Cultivars: Parbhani Moti, Phule Chitra, Phule suchitra, Kesalpoor village seed, 1004

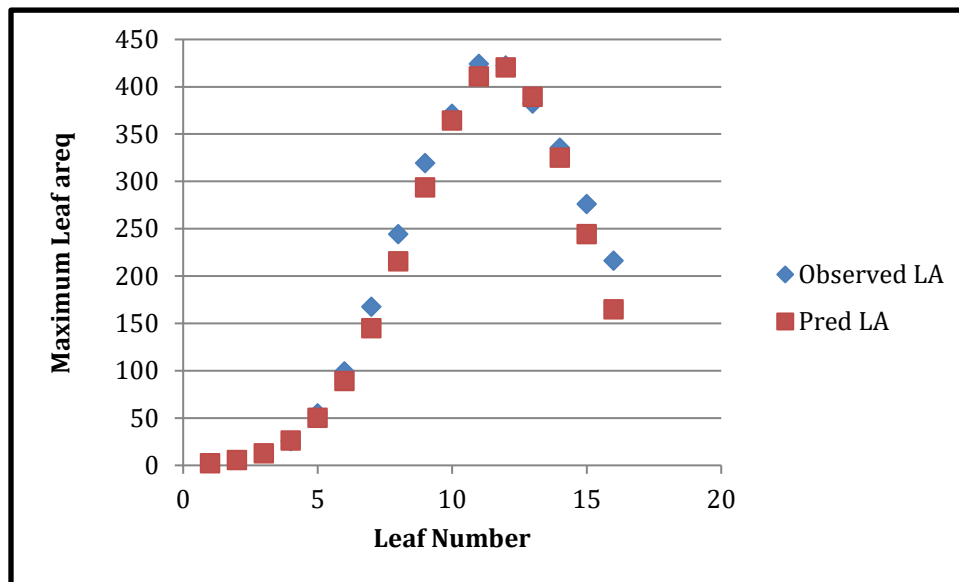


Fig 8d. Optimization for ILA function for cultivars Parbhani Moti, Phule Chitra, Phule suchitra, Kesalpoor village seed, 1004

Repetition of field experiments:

The field experiments were repeated in the second year (2018-19 Rabi season) with the same treatments. The data collected in the first and second year will be compiled and then used for parameterization and validation of APSIM.

Yield Validation with NASA data:

Comparison of yield and biomass by running model with observed weather data versus running model with gridded NASA data. We found that simulated yields (both grain yield and biomass) with NASA data were in better agreement with simulated yields with observed weather data.

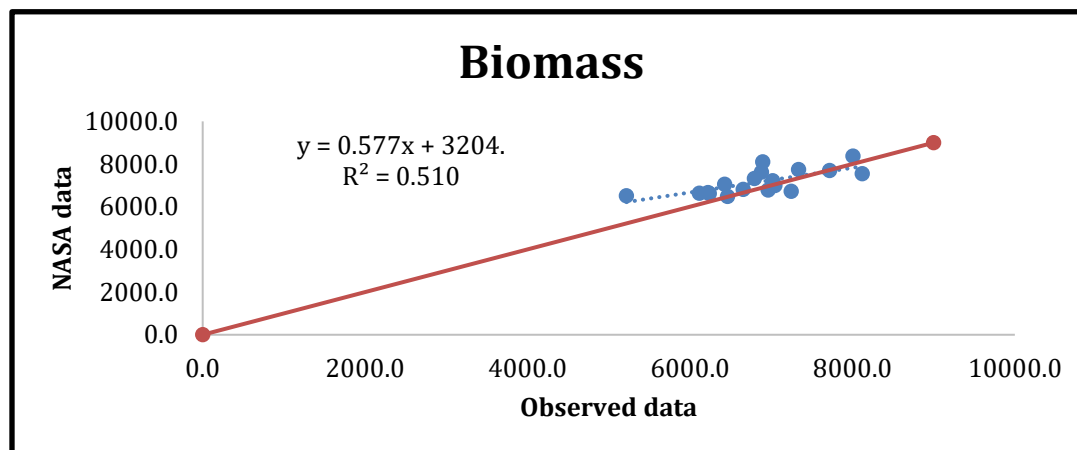


Figure 9 - Comparison of biomass by running model with observed weather data versus running model with gridded data of NASA

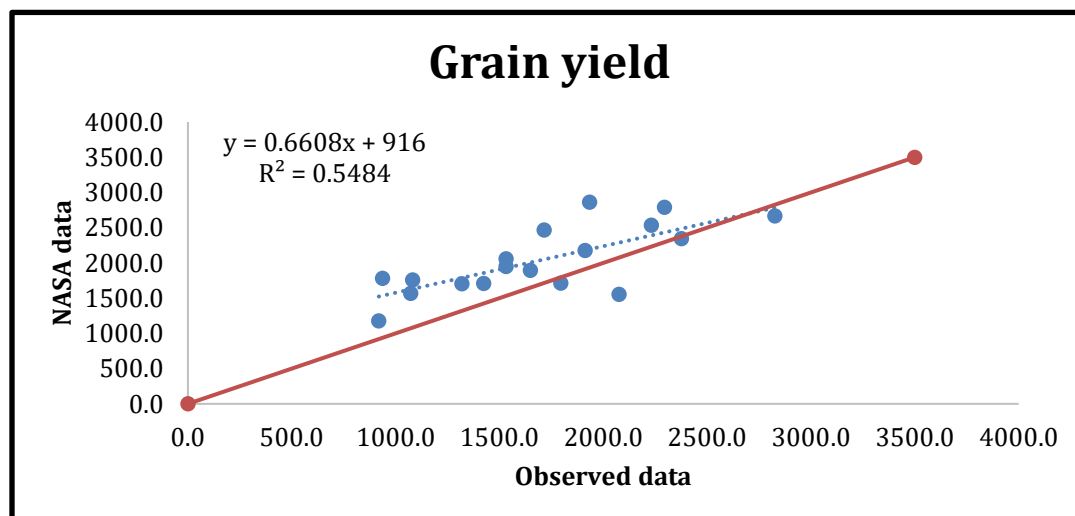


Figure 10 - Comparison of yield by running model with observed weather data versus running model with gridded data of NASA

Annex 3

Progress report on the project “Determination of bean breeding targets in Colombia”

Principal investigator: Julian Ramirez-Villegas (CIAT)

Lead partner: Andy Challinor (University of Leeds, UK)

Project contributors: Jeison Mesa (CIAT), Patricia Alvarez (CIAT), Chetan Deva (University of Leeds), Ioannis Droutsas (University of Leeds), Mikolaj Swiderski (University of Leeds).

The present report summarises preliminary results for the Mini-grant “*Determination of bean breeding targets in Colombia*” within the CGIAR Platform for Big Data in Agriculture. The project aims at identifying stresses and selection sites for breeding bean for abiotic stress in Colombia using crop-climate modelling. The project leverages the BBSRC-funded project “Bean breeding and Adoption in changing climates in post-conflict Colombia (BACO)”. Project deliverables are as follows,

- D1. Month 3. Analysis strategy including representative cultivars for which field data collection and modelling analysis will be conducted, and crop model to be used.
- D2. Month 8. Calibrated model for Colombian conditions using CIAT data.
- D3. Month 10. Climate change-determined abiotic stress patterns based on model runs.
- D4. Month 12. List of candidate sites for selection produced and discussed with bean breeders.

The previous report (4th Sept 2018) summarized data gathering and initial calibration of the SAB686 variety. Progress reported here relates to the completion of deliverable D2 (model calibration), with preliminary results also presented for deliverable D3 (analysis of stress patterns). Finally, a summary bulleted list of next steps is presented.

Model calibration and evaluation

The CSM-CROPGRO-DRYBEAN model was calibrated and evaluated using detailed growth and development data from an experiment conducted in 2014 in the locality of Villanueva in the Santander department in Colombia. The model was calibrated and evaluated for two varieties included in the trial, a representative commercial check (Calima) and a promising line for drought and heat stress for Colombia (SAB686). Genotype SEF60 has not been included in this work due to lack of data. These two varieties were deemed representative enough of

commercial and breeding genotypes, and will hence be used for all analyses. The field experiment included seven treatments (i.e. seven different planting dates), out of which two were used for model calibration, and the remaining four were used for model evaluation.

Model calibration was performed through the following steps:

- Identify key parameters that needed calibration through sensitivity analysis. To this aim, we conducted a Sobol (Sobol et al., 2007; Saltelli et al., 2007) global sensitivity analysis for both growth and development parameters in the crop model. We conducted this analysis for three representative sites in Colombia, namely, Palmira (3° 35' 0" N, 76° 15' 0" W), Santander de Quilichao (3°01' 26.72" N and 76° 29' 15.67" W), and Barichara (6° 38' 10" N and 73° 13' 25" W). Phenology parameters included in the analyses were the photothermal time (PTT) between emergence and flower appearance (EM-FL), PTT between first seed and physiological maturity (SD-PM), and PTT between first flower and first seed (FL-SD). Growth parameters included the specific leaf area under standard growth conditions (SLAVR), maximum size of full leaf (SIZLF), maximum leaf photosynthesis rate (LFMAX), and the maximum weight per seed (WTPSD).
- Calibration of model parameters through a genetic algorithm with calibration data (2 treatments). Using observed dates of flowering, pod and seed appearance and physiological maturity, we first calibrated phenology parameters. Next, we calibrated growth parameters using total, leaf, stem, and pod biomass, yield, and leaf area index. For phenology, for which all variables are in the same units (i.e. days), we sought to minimize the sum squared error (SSE), whereas for growth, we used the distance correlation normalized by the mean absolute error.
- Evaluation of the calibrated model with evaluation data (4 treatments). For SAB686 we also tested the model against a number of CIAT experiments in other localities than Villanueva. For Calima, we compared our calibration results with those of Clavijo-Michelangeli et al. (2014) (CM14 calibration hereafter). For more details on the Clavijo-Michelangeli et al. (2014) experiments and data the reader is referred to Hwang et al. (2017).

Fig. 1 shows phenology calibration and evaluation results for both SAB686, Calima (our calibration) and Calima (CM14 calibration). In general, the simulation of phenology is found to be accurate for both calibration and evaluation experiments, with RMSE being the smallest for SAB686 (RMSE=2.47 days, RMSE relative to mean = 4.8 %), and about twice as large for Calima (4.95 days, 9.8 %). Notably, RMSE for both reduces in the evaluation experiments (except for CM14), suggesting the model represents well the phenology of these genotypes. We also note that CM14 calibration performed better than the calibration reported here in both calibration and evaluation experiments.

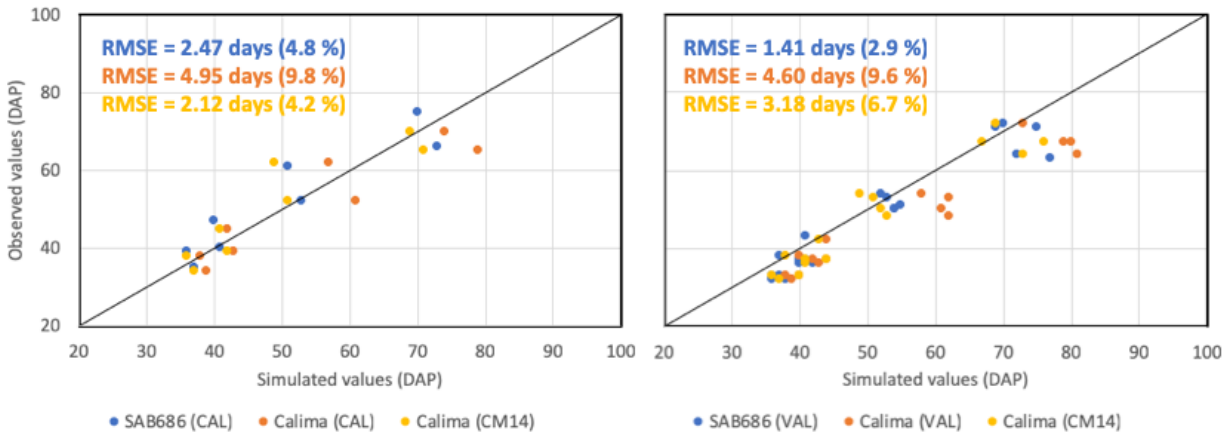
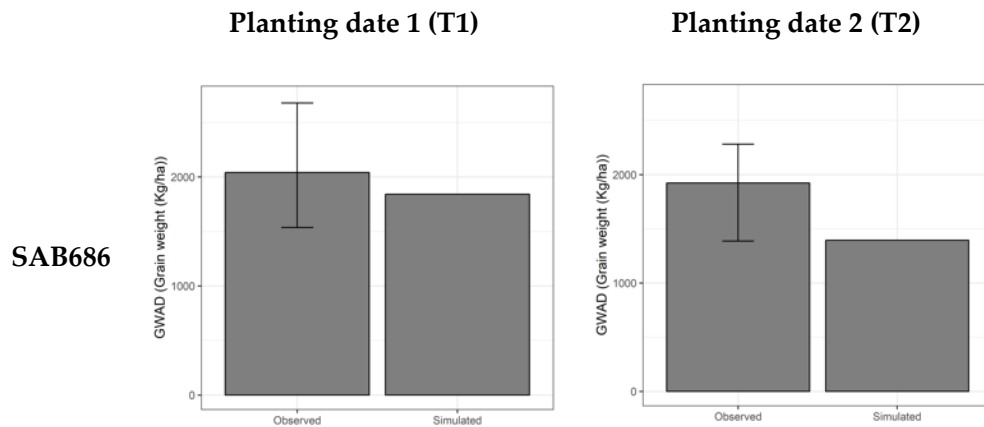


Figure 1 Calibration (left) and evaluation (right) results for genotypes SAB686 and Calima

Simulated yield is presented in Fig. 2 for the calibration experiments. For these experiments, in general, the performance of the model for simulating growth (not shown) and yield of SAB686 was better than for Calima. For Calima, specifically, we note that CM14 calibration performed poorly compared to the calibration reported here. There were also differences in how well the model simulated both treatments across the two genotypes. Treatment 1 was systematically well simulated for both genotypes, whereas treatment 2 was generally underestimated by the model, especially for Calima.



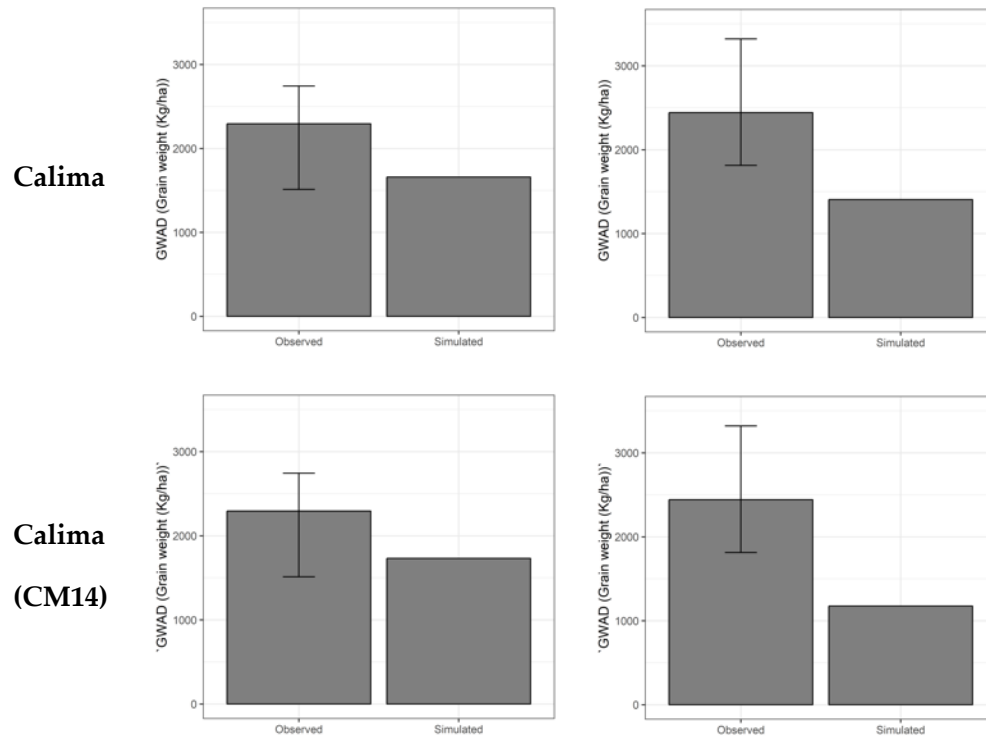


Figure 2 Simulated yield for calibration experiments for genotypes SAB686 and Calima. Results for the parameters reported by CM14 are also shown.

Simulated yield for evaluation experiments is shown in Fig. 3. We find consistent model performance across the three genotypes, with the model generally performing well for treatment 3 and 5, and relatively poorly in treatment 4. It is possible that the model is over-estimating the degree of drought stress in treatment 4 for the three genotypes, or that there is associated experimental error in the field measurements, especially for genotype Calima, for which variation across replicates is larger. It is also possible that the weather measurements or the initial field conditions are not captured well. As for calibration experiments, the performance of the model for SAB686 was greater compared with that of Calima. Finally, we note that the CM14 growth parameter set tends to perform less well than our calibration. These results suggest that there are inherent limitations in the model that require more careful examination and improvement, especially under stress situations and for genotype Calima. However, despite these limitations, the model performs sufficiently well for phenology, growth and yield, especially for genotype SAB686.

Planting date 3 (T3)

Planting date 4 (T4)

Planting date 5 (T5)

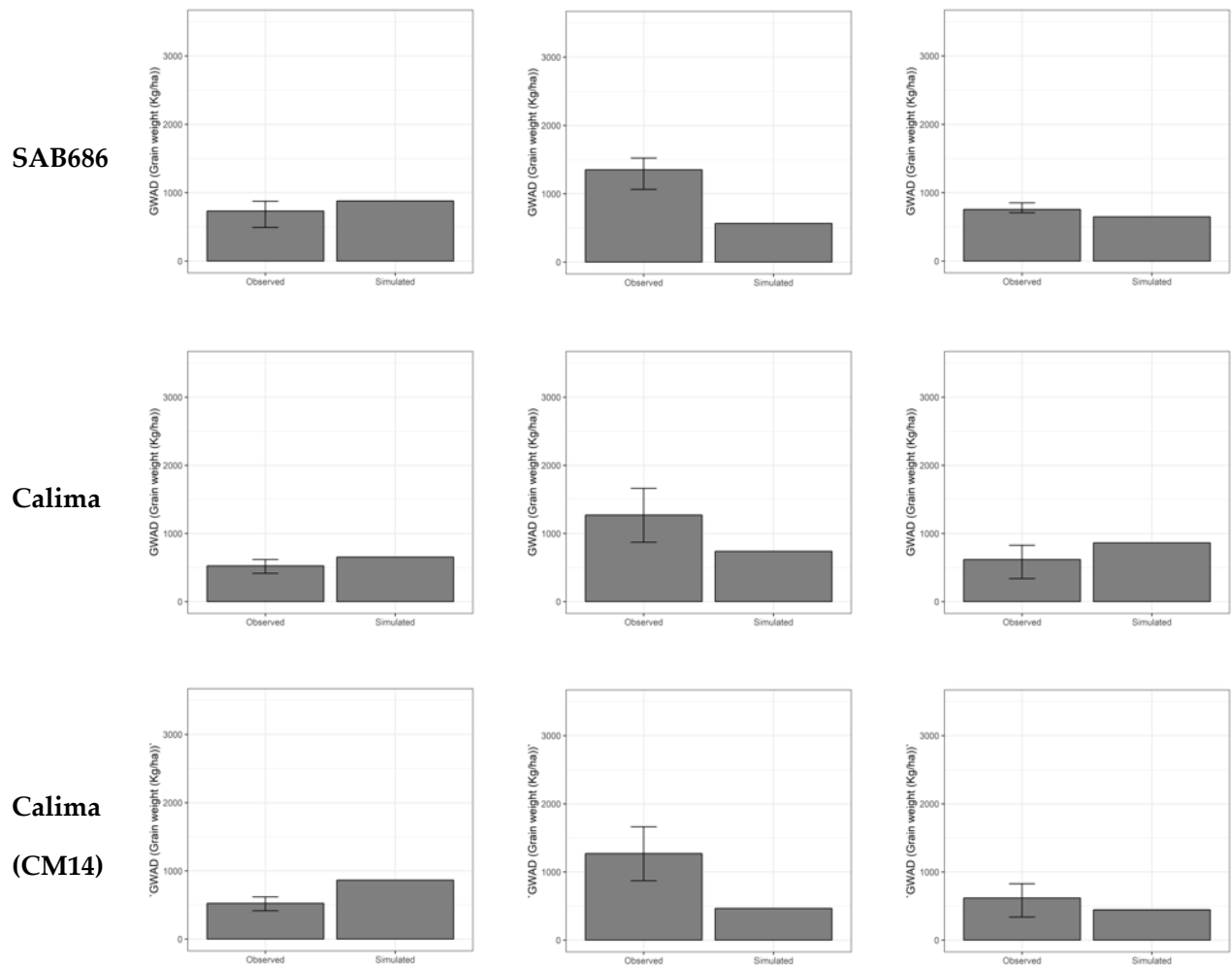


Figure 3 Simulated yield for evaluation experiments for genotypes SAB686 and Calima. Results for the parameters reported by CM14 are also shown. Only three out of the five evaluation treatments are shown.

Model parameters are shown in Table 1 for the two genotypes. As a reference, we include parameters from CM14, which are generally consistent with the calibration results from the genetic algorithm.

Climate change-determined abiotic stress patterns

With the calibrated model, we simulated crop growth of the common bean initially in a set of five representative sites in Colombia. Two of the five sites have temperatures near the optimal for common bean, namely, Palmira and Santander de Quilichao. One site (Barichara) is sub-

optimal in terms of temperature. Lastly, two sites have supra-optimal temperatures during the bean growing season (Espinal and La Uribe). Simulations were conducted for both current and future climate conditions. To represent current climates, we used observed 1981–2010 climate data, whereas for future climates we used 2040–2069 projections of the GFDL-ESM2 General Circulation Model (GCM) under RCP 8.5. Future climate data were bias-corrected using the change factor method (Hawkins et al., 2013). Simulations were run for a range of planting dates and only for one of the genotypes (SAB686).

Table 1 Calibrated model parameters for genotypes SAB686 and Calima

Parameter	Description	Units	SAB686	Calima	Calima (CM14)
ECO#	Ecotype code	–	ANDEET	ANDEET	ANDEET
CDSL	Critical daylength for photoperiod sensitivity	hour	12.170	12.170	12.500
PPSEN	Slope of the relative response of development to photoperiod with time	1/hour	0.00	0.00	0.11
EM-FL	Time between plant emergence and flower appearance	PT days ¹	24.00	25.70	24.10
FL-SH	Time between first flower and first pod	PT days	3.00	3.00	3.50
FL-SD	Time between first flower and first seed	PT days	11.90	14.90	10.00
SD-PM	Time between first seed and physiological maturity	PT days	17.78	16.11	19.50
FL-LF	Time between first flower and end of leaf expansion	PT days	10.00	10.00	11.00
LFMAX	Maximum leaf photosynthesis rate at 30 C, 350 vpm CO ₂ , and high light	mg CO ₂ /m ² s	1.11	1.07	0.94
SLAVR	Specific leaf area of cultivar under standard growth conditions	cm ² /g	271.00	276.00	230.00
SIZLF	Maximum size of full leaf (three leaflets)	cm ²	133.80	151.30	300.00

XFRT	Maximum fraction of daily growth that is partitioned to seed + shell	–	1.00	1.00	1.00
WTPSD	Maximum weight per seed	g	0.972	0.960	0.450
SFDUR	Seed filling duration for pod cohort at standard growth conditions	PT days	15.00	15.00	16.00
SDPDV	Average seed per pod under standard growing conditions	#/pod	3.50	3.50	4.00
PODUR	Time required for cultivar to reach final pod load under optimal conditions	PT days	10.00	10.00	5.00
THRSH	The maximum ratio of (seed/(seed+shell)) at maturity.	–	78.00	78.00	75.00
SDPRO	Fraction protein in seeds (g(protein)/g(seed))	–	0.235	0.235	0.235
SDLIP	Fraction oil in seeds (g(oil)/g(seed))	–	0.03	0.03	0.03

¹ Photothermal days

Using these simulations, we conducted a preliminary Target Population of Environments (TPE) analysis, with particular focus on temperature response. We classified the five sites and all simulated years into three TPEs: cool, warm, and optimal. In the current climate (Fig. 2 - solid lines), the optimal TPE (green solid line) has considerably more high yield simulations than the colder (blue solid line) and hotter (red solid line) TPEs. By 2050 (dashed lines), the difference between temperature TPEs narrows: low yield seasons become more common in the above optimal TPE, closing the gap in simulated yield between the hotter and colder environments (Fig. 2 - dashed lines). Notably, the future projected yield of the optimal TPE becomes increasingly similar to that of the warm TPE, possibly suggesting similar patterns of stress. This will be a matter of further investigation.

These results indicate that TPEs useful means of capturing changes in bean stress patterns. The difference in yield between the wet (optimal) and dry TPEs also narrows (not shown), as both

conditions experience a greater frequency of poor harvests. Taken together, these results suggest that changes in temperature, not rainfall will determine changes in TPEs.

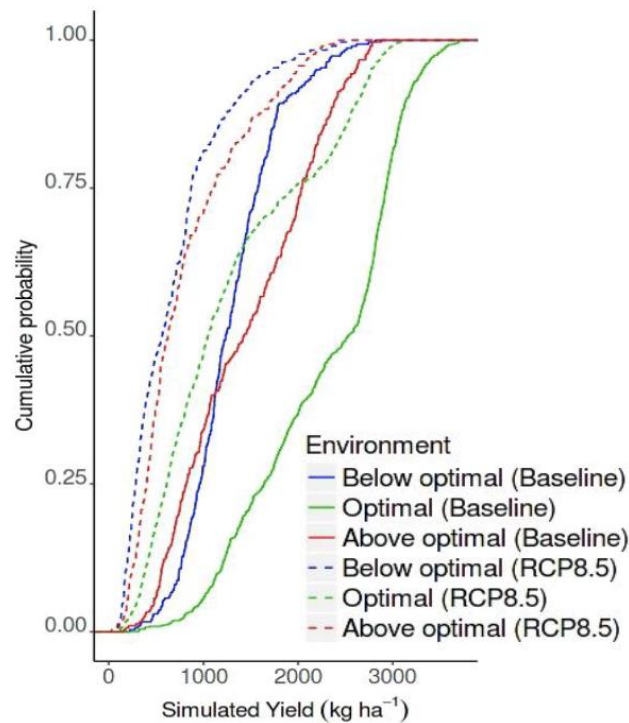


Figure 4 Temperature TPEs - current vs. 2050 using the DSSAT CROPGRO-DRY BEAN model. Climate projections from the GFDL-ESM2 models.

Conclusions

Several conclusions become apparent with the analyses presented here. Firstly, the model simulates well the overall phenological and growth behavior of the crop. However, due to the small variation in the dates of anthesis, first pod and seed, and physiological maturity, it is difficult to tell whether the model's representation of individual phenological events is correct. A similar issue is found for growth and yield simulation. We deem necessary the testing of the model with additional field data. We, however, conclude that the model can be used to investigate TPEs in Colombia, especially for genotype SAB686. A revision of results and parameters for Calima will be needed in order to improve simulation results.

Our analysis of temperature TPEs indicates distinct responses for the different temperature environments analyzed. Since the future projected yield of the optimal TPE becomes

increasingly similar to that of the warm TPE, it is possible that current optimal growing areas in Colombia are likely to become heat stressed, thus requiring focusing breeding efforts on heat stress adaptation. Representative sites toward this aim would be Espinal (Tolima) and La Uribe (Meta).

Next steps

- Scale the TPE analysis to a broader set of sites, including other temperature regimes and post-conflict areas with potential to grow beans.
- Revise calibration for Calima, and perform model simulations and TPE analyses in all selected sites with the two calibrated genotypes.
- Discuss results with bean program experts and propose target selection sites for bean breeding for heat stress.

References

- Clavijo Michelangeli, J.A.; Boote, Kennet J.; JONES, J.W.; Corell, M.; Gezan, S.; Bhakta, M.; Zhang, L.; Osorno, J.; Rao, I. M.; Beebe, S. E.; Roman-Paoli, E.O.; Gonzalez, A.; Beaver, J.; Ricaurte, J.; Colbert, R.; Carvalho, M.; Vallejos, C.E.. 2014. Modeling genetic traits of five common bean (*Phaseolus vulgaris*) genotypes in multi- location trials . University of Florida, Florida, FL, USA. 1 p. (Poster presented at ASA, CSSA, & SSSA International Annual Meeting. Nov. 2-5, 2014, Long Beach. CA.
- Hawkins E, Osborne TM, Ho CK, Challinor AJ (2013) Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agric For Meteorol* 170:19–31
- Hwang C, Correll MJ, Gezan SA, et al (2017) Next generation crop models: A modular approach to model early vegetative and reproductive development of the common bean (*Phaseolus vulgaris* L). *Agric Syst* 155:225–239 . doi: 10.1016/j.agsy.2016.10.010
- Saltelli A, Ratto M, Andres T, et al (2007) *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, Chichester, UK.
- Sobol' IM, Tarantola S, Gatelli D, et al (2007) Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliab Eng Syst Saf* 92:957–960 . doi: 10.1016/j.ress.2006.07.001

Annex 4

Combining crop and disease modeling with numerical weather forecasting to inform wheat blast early warning systems in Bangladesh and Brazil

Progress report number 1 (July – December 2018)

José Mauricio Cunha Fernandes¹, Timothy J. Krupnik^{2*}, Felipe De Vargas³, Akbar Hossain^{4*}, Sk. Ghulam Hussain.

1. Plant pathologist. The University of Passo Fundo, Av. Brazil Leste, 285 - São José, Passo Fundo - RS, 99052-900. Brazil. Email: jmauricio@upf.br

2. Senior Scientist and Systems Agronomist. International Maize and Wheat Improvement Center (CIMMYT) | Sustainable Intensification Program. House 10/B. Road 53. Gulshan-2. Dhaka, 1213, Bangladesh. Email: t.krupnik@cgiar.org

3. Computer Scientist. Av. Brazil 565 – Apto. 602. Passo Fundo, RS, Brazil. 99010-000. Email: felipevar26@gmail.com

4. Senior Agronomist. Bangladesh Wheat and Maize Research Institute, Nashipur, Dinajpur-5200, Bangladesh. Email: akbarhossainwrc@gmail.com

5. Sk. Ghulam Hussain. CIMMYT. House 10/B. Road 53. Gulshan-2. Dhaka, 1213. Bangladesh. Email: g.hussain@cgiar.org

*Corresponding author email: t.krupnik@cgiar.org

Correct citation:

Fernandes, J.C.M, Krupnik, T.J. De Vargas, F., Hossain, A., Ghulam Hussain, Sk. 2019. Combining crop and disease modeling with numerical weather forecasting to inform wheat blast early warning systems in Bangladesh and Brazil. Progress report 1 (July – December 2018). International Maize and Wheat Improvement Center, Dhaka, Bangladesh.

Data availability:

Weather and crop data utilized in simulations are available online in open-source format at <http://dev.sisalert.com.br/shiny/cropwheat/>.

Summary ‘in a nutshell’

Wheat blast, caused by *Magnaporthe oryzae* pathotype *Triticum* (MoT), is a potentially severe disease of wheat. Wheat blast infections vary with prevailing climatic conditions, the degree of susceptibility of host cultivars, and the location of infection. Most programs working to mitigate the threat of disease infection focus on host resistance and/or calendar based fungicide application. The latter is a particular concern where environmental and human health is a concern; fungicides are also costly and smallholder wheat farmers may not be able to afford regular calendar based application. An alternative approach is to use weather forecasts and disease models when and where disease might occur, with preventative management advisories delivered to farmers several days in advance of potential disease outbreaks. This report details modeling efforts to develop such an ‘early warning system’ (EWS) in Bangladesh using insights and models originally developed in Brazil. In addition, novel work to integrate the DSSAT-Nwheat crop production simulation model with a weather-data driven blast disease development model is presented. Although this research is very preliminary and results contained within this report should be treated very conservatively, we document the first-ever effort to dynamically simulate and couple both the MoT life cycle and wheat crop development by linking weather-data driven disease and crop models to estimate infection and spike damage. Further research that builds on these preliminary efforts will be incorporated into a numerical-weather model driven and geographically explicit EWS for wheat blast under development in Bangladesh in Brazil.

Keywords

Simulation model coupling, Decision Support System for Agrotechnology Transfer (DSSAT), wheat blast disease (*Magnaporthe oryzae* pathotype *Triticum* (MoT)), wheat (*Triticum aestivum*), early warning system.

1. Introduction

Wheat blast, caused by *Magnaporthe oryzae* pathotype *Triticum* (MoT), was first discovered in six municipalities in the Parana state of Brazil in 1985 (Igarashi et al., 1986). While the disease attacks all above ground plant parts, fungal infection of wheat spikes by MoT can cause significant yield losses (Cruz and Valent, 2017). While spike infection by MoT resembles *Fusarium* head blight, the two diseases can be differentiated because MoT infects the rachis, above which spikes appear white with partially or completely unfilled grain. Black spotting of the entire spike can be seen with the naked eye in advanced infections. Yield loss is reported to be most severe when wheat is infected during the flowering and early grain formation stages.

Following initial appearance of the disease in Brazil, wheat blast spread to Bolivia, Paraguay and parts of Argentina (Duveiller et al., 2016). In 2016, MoT unexpectedly appeared in South Asia, causing losses on >15,000 ha in Bangladesh (Malaker et al., 2016). At that time, wheat was Bangladesh's second most widely grown cereal crop. Large-scale infection sparked fears of potential spread across the region. The disease reappeared in Bangladesh during the 2016/17 and 2017/18 wheat growing seasons, with reports of spread to India appearing in the popular press (e.g. Das 2017). Although prevailing climatic conditions and a significant reduction in wheat areas in Bangladesh limited the area of infection in these seasons, the threat posed by wheat blast remains very real. Regional extrapolation estimates have suggested that South Asian farmers could lose up to 1.77 million tons year⁻¹ with light infections of just 10% (Mottaleb et al., 2018).

Wheat blast infections vary with prevailing climatic conditions, the degree of susceptibility of host cultivars, and the location of infection (leaf, stem or spike) (Goulart et al., 2003). Fungicides are ineffective under high disease pressure and partially effective under moderate-to-low pressure. A 2NS translocation segment from *Aegilops ventricosa* was reported to confer moderate resistance to wheat blast in some genotypes (Cruz et al. 2016). However, in Bolivia, during the 2015 epidemic, the best resistance available was insufficient for controlling blast (Vales et al. 2018). In Bangladesh, a resistant variety (BARI Gom 33 which carries the 2NS translocation segment) is now publicly available, although seed is in limited supply and the durability of resistance remains unclear.

Despite this information, most programs in Bangladesh focus on advising farmers on how to reduce wheat blast risks concentrate on seed treatment and calendar based fungicide application. This approach can be costly and relies on assumptions of large-scale fungicide availability. An alternative approach to mitigating the threat posed by wheat blast is to incorporate principles of integrated pest management (IPM), although most IPM programs require farmers to regularly visit and observe their fields to estimate infection levels. In addition to the time-consuming nature of repetitive field visits, this approach is arguably problematic because it requires an advanced knowledge of the disease symptoms. Yet once symptoms are visible, infection is already underway. This limits the farmers' to preventatively control the disease with fungicides.

An alternative approach that can be used as part of an IPM strategy is to use weather forecasts and disease models when and where disease might occur, with preventative management advisories delivered to farmers several days in advance of potential disease outbreaks. Such 'early warning systems' however require knowledge of disease ecology, reliable weather forecasts and significant efforts to properly model disease outbreaks before advising farmers.

MoT tends to be associated with regions that have high humidity (Fernandes et al. 2017). Combined with temperature, relative humidity influences the speed of *M. oryzae* development (Calvero et al. 1996). Cardoso et al. (2008) suggested that the optimal temperatures for blast infection of wet or damp spikes range between 25 and 30 °C. Spore development of MoT is similar to *Magnaporthe oryzae* pathotype *Oryzae* (MoO), which infects rice. Bregaglio and Donatelli (2015) developed a predictive model for MoO inoculum development within the cropping season. The model simulated the pace of conidiophore development as a function of hourly temperature and relative humidity. Fernandes et al. (2017) subsequently adapted this model to MoT using data from Brazil and simulated the development of blast spore clouds to indicate dates when wheat infection might occur. Unlike other models that require a considerable amount of input information, Fernandes et al. (2017) sought to simplify input parameters to increase the ease by which users can make use of the model. Several parameters common in plant disease models, for example duration of leaf wetness, are not included in the empirically based modeling approach.

To assess the potential impacts of pests and diseases on crop production, some crop models include coupling points, which can be thought of as special model variables whose changing values can be used to represent pest and disease damage to crop organs or growth processes. The coupling point concept was first introduced in 1983 (Boote et al. 1983), and later formally implemented in the DSSAT crop modeling platform (Jones, et al., 2003) as a Pest Module (Batchelor et al., 1993). Examples of coupling point variables include leaf mass or area, stem mass, root mass, root length, and seed mass or number, all of which might be negatively impacted by pests and diseases. By identifying specific mechanical and infection damage pathways, and rates of damage using these variables, growth models can be tuned to quantify how crop development and yield might be affected. The DSSAT Pest Module allows users to input field observations or dynamically modeled data on insect damage, disease severity, and physical damage to plants or plant components (e.g., grains or leaves). Users can then simulate the likely effects of those pests and diseases on crop growth and economic yield.

This report provides information on initial progress to dynamically simulate and couple both the MoT life cycle and wheat crop development by linking weather-data driven disease and crop models to estimate infection and spike damage. Although this project focuses on both Bangladesh and Brazil, this report provides details of initial model coupling and preliminary simulation efforts in Bangladesh only. Simulations for Brazil and for multiple locations in Bangladesh will be presented in the final project report in the third quarter of 2019. All results presented in this report should therefore be treated as preliminary and non-definitive.

2. Materials and Methods

2.1. Model background

The Nwheat model simulates wheat growth under differing environmental and management conditions. The model allows users to compare simulation results to potential (unconstrained) growth. Nwheat was recently embedded in the Decision Support System for Agrotechnology Transfer (DSSAT, Version 4.7) model as an Agricultural Production Systems sIMulator (APSIM) Nwheat model (Kassie et al., 2016). This model has been tested under the shell of APSIM-Nwheat in a number of environments variable in temperature, carbon dioxide, nitrogen and soil-water conditions. Conveniently, DSSAT-Nwheat uses the same input information as CERES-Wheat. Additional cultivar genetic coefficients are nonetheless needed during calibration.

DSSAT-Nwheat is currently under developer improvement to enable simulations that account for biotic stresses to crop growth, most notably for the damage imposed by diseases and pests (Farman et al., 2017). Lazzaretti et al. (2016) parameterized a generic disease simulator to mimic wheat blast inoculum build-up and consequent colonization of foliage and wheat spikes. An inverse relationship between wheat grain weight without disease control and the spread of fungal mycelia within the spikes was programmed into the model, representing the effect of increasing wheat blast intensity on the crop. This routine was coupled with a weather-based wheat blast simulation developed in Brazil by Fernandes et al. (2017) that requires hourly observations of temperature, relative humidity, rainfall, and solar radiation. The coupling of this model to the DSSAT-Nwheat is described in this report. Our work to couple the models is novel and is intended to predict yield in the presence or absence of wheat blast.

To achieve these aims, non-relational open source database known as MongoDB was utilized to implement a flexible structured data schema. This permitted the storage and retrieval of experimental data and model-required input data. The MongoDB is shown in Figure 1. Data utilized in simulations are available online in open-source format at <http://dev.sisalert.com.br/shiny/cropwheat/> by clicking 'data explorer' for any chosen simulation. Preliminary efforts to facilitate model coupling and calibrate this model in Bangladesh are described in this report.

2.2. Experimental design, management and data collection

Field experiments to calibrate model performance were carried out in Bangladesh in the 2017/18 wheat growing season. A split-plot design six wheat cultivars and five sowing dates was utilized. The experiment was replicated in three locations in Bangladesh. This focuses research locations managed by the Bangladesh Maize and Wheat Improvement Institute (BMWII) in Dinajpur (25°44'38.62"N, 88°40'21.36"E) and Jessore (23°11'11.71"N, 89°11'16.06"E). The former location was used for model calibration, the latter location was used for virtual experimentation.

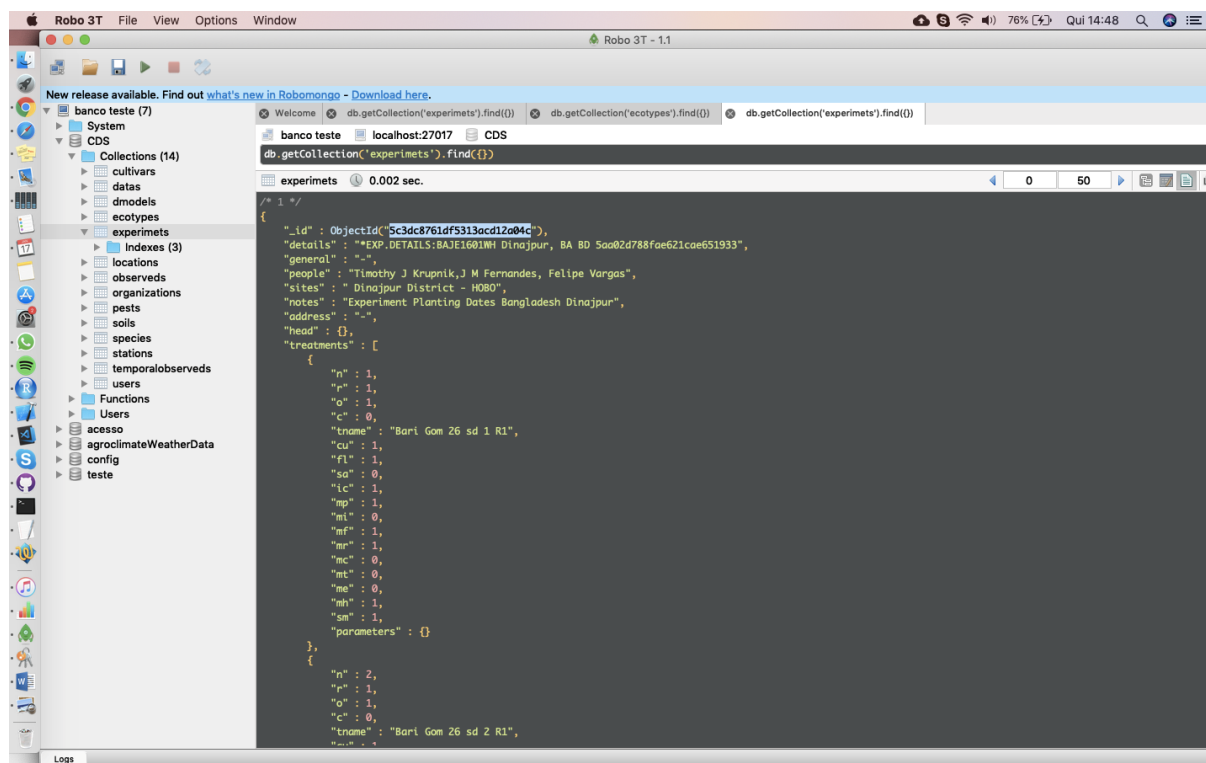


Figure 1. DSSAT-Nwheat FileX in text and json format used in the context of a non-relational open source database known to implement a flexible structured data schema.

Experimental sub-plot size was 6m² (3m row length and 10 rows spaced 20 cm apart). Sowing dates included November 25, December 2, December 12, December 22 and January 1. Cultivars tested included BARI Gom 26, 28, 30, 31, 32 and 33. Seed was sown by hand at between 5-7 cm depth following two tillage passes and covered with soil. Seed rates were 120 kg ha⁻¹ for all varieties exempting BARI Gom 33 which tillers poorly, for which rates were increased by 20 kg ha⁻¹. This report focuses on performance and calibration for BARI Gom 26. Based on field observations, this cultivar is classified as susceptible to blast. Other varieties will be considered in the next report.

Fertilizers were applied at elemental rates of 100-27-40-20-1 kg ha⁻¹ of N-P-K-S-B. Two-thirds of N and the full amount of the other fertilizers were applied basally. The remaining N split was applied immediately before the first irrigation near crown root initiation (17-21 days after sowing (DAS)). Three light irrigations no more than 5 cm deep were applied. The first was applied as described above, the second and third irrigations were applied at booting (50-55 DAS) and grain-filling (70-75 DAS) stages. Weeds were controlled to reduce any competition with the crop.

After excluding 20 cm borders to minimize edge effects, the remaining plot was harvested at physiological maturity a corrected to a moisture content of 12%. Measured crop variables from experiments included anthesis and maturity dates, total above ground biomass, and grain yield. 1,000 grain weight was also measured.

2.3. Meteorological data

The DSSAT-Nwheat model combination requires daily precipitation, maximum and minimum air temperature, and solar radiation as standard weather inputs. In Bangladesh, these inputs were taken from HOBO automated weather stations (Bourne, MA, USA) placed at 2 m height from less than 200 m from the experiments. Any gaps in data due to equipment malfunction were filled with data from the nearest Bangladesh Department of Meteorology (BMD) synoptic observation station. The wheat blast simulator required hourly temperature and relative humidity data from the same sources. Where only three-hourly data were available (as was the case with BMD stations), data were interpolated to obtain hourly values.

2.4. Description of the wheat blast environmental suitability modeling system

A wheat blast simulation model that accounts for inoculum build-up and infection has been validated in Brazil (Fernandes et. al., 2017). This model forms the basis disease severity estimates this preliminary research coupled with the DSSAT-Nwheat model to simulate disease-induced yield loss. Fernandes et. al. (2017) developed and evaluated a prediction model based on the analysis of historical epidemics and weather series data in the northern Paraná state, Brazil. Local available epidemiological knowledge was also employed in model parameterization. Importantly the model also assumes the spatially uniform presence of MoT inoculum in the environment for which simulations are run.

The disease and hourly-scale weather datasets examined by Fernandes et. al. (2017) for Brazil encompassed the 2001–2012 period. A specific database management application developed using R Shiny was programmed to visualize and identify patterns in weather variables during two major outbreaks (2004 and 2009). Uncommonly humid and warm weather was observed for most locations in this study during a 60-day period preceding wheat heading during years of major outbreaks. These conditions were therefore considered key drivers of inoculum build-up and airborne spores from regional inoculum sources in the surroundings.

The blast model developed by Fernandes et. al. (2017) has four components. The first component assumes spores are present and estimates the rate of conidiophore development as a function of temperature and relative humidity, both of which are integrated to estimate blast inoculum potential by solving equation 1 for the hourly sum of inoculum potential (IP) over the season:

$$IP = \begin{cases} 14.35 - 0.25 * T & \text{if } 15^{\circ}C < T < 27^{\circ}C \text{ and } RH \geq 93\% \\ -8.5 - 0.59 * T & \text{if } 27^{\circ}C < T < 35^{\circ}C \text{ and } RH \geq 93\% \\ 0 & \text{otherwise} \end{cases} \quad [1]$$

where IT and RH are air temperature and relative humidity, respectively. Where RH is below the threshold in Equation 1, the model does not accumulate thermal time. The model also

calculates the development of a spore cloud subject to assumptions of air current uptake, atmospheric diffusion and wind shear that affect spore longevity. Survival of spores while airborne may also be affected by temperature, solar and ultraviolet radiation, in addition to relative humidity (Deacon 2005). Spore cohorts were therefore assumed to have a half-life of three days within any seven-day window.

The model also determines the number and timing of days with climatic conditions favoring blast infection using a conditional ruleset. Days favoring infection were consequently declared following spore cloud development when the daily maximum temperature exceeded 23°C and temperature amplitude (calculated daily minimum temperature subtracted from daily maximum temperature) was > 13°C, with mean daily RH above 70%. The model adequately described observed epidemic and non-epidemics years during and beyond the study period in Brazil. This report details preliminary efforts to transfer this model to Bangladesh while also coupling it with DSSAT-Nwheat.

2.5. Integration of crop and disease models

Because a pest simulation module in DSSAT has been made accessible to the DSSAT-Nwheat model, a communication approach was used to couple the parameterized wheat blast model to DSSAT-Nwheat. Using this system, models are executed as independent programs that can communicate with each other during execution using a Message Passing Interface (MPI). This effectively permits the passing of information between models and allows automated changes in the results produced by one model based on the outcomes of the other (Browne and Wilson, 2015).

When applied in Bangladesh, the genetic coefficients of the model were refined to numerically represent the genotypic constitution of wheat cultivar BARI Gom 26. These genetic coefficients express all relevant information regarding the life cycle, distribution and duration of each phenological stage, growth rates indexes, biomass distribution and reproductive processes. The genetic coefficients BARI Gom 26 are presented in Annex 1.

In this study, the development of a customized graphical user interface (GUI) was motivated by the need to enhance researchers' capability to manipulate DSSAT-Crop Simulation Model (CSM) files to achieve simulation and data visualization goals. As such, an interactive library that could provide a mechanism for reading, writing and processing different DSSAT-CSM files was needed (Figure 2). There was also a need to implement software engineering best practices to achieve cross-platform goals. Windows, MacOS X and Linux are three of the world's most common operating systems. A JavaScript library was consequently created to read, write and process DSSAT-CSM files using DSSAT data standards with each of these operating systems (White et al., 2013). The library is also able to integrate with different output files (.OUT extension) from the DSSAT-CSM, for example plant growth, soil water, weather information, and other .OUT files. Each output file is generated with meta-data including information about the model, experimental runs, treatments and simulated values.



Figure 2. DSSAT-Nwheat FileX in text and json format.

The GUI was created using R Shiny to visually represent model coupling and allow users to point and click and run virtual experiments. The beta form of this GUI can be found at: <http://dev.sisalert.com.br/shiny/cropwheat/>. The overall concept for the integration of both crop and disease models as important components of a numerical weather model driven wheat blast early warning system can be found in Figure 3.

The DSSAT-Nwheat application uses the MongoDB database to store outputs from simulations and collates these alongside observed data. The R software embedded in the application enables the development of more robust and flexible graphs within the user interface. This in turn allows GUI users to produce graphs of simulated outputs with observed data and associated statistics.

2.6. Virtual experimentation and data analysis

The goal of this research was to develop a proof of concept that models could be coupled to simulate wheat blast impact on yield in Bangladesh. To this end, we calibrated the DSSAT-Nwheat CSM by comparing it to experimental results for grain and biomass yield, 1,000 grain weight, and phenology. DSSAT-Nwheat model calibration was achieved by manipulating the genetic coefficients of each cultivar by iterative trial and error until observed data sufficiently matched simulated data.

We implemented three virtual experiments for the wheat cultivar BARI Gom 26 using the customized GUI and simulated wheat productivity performance with and without the blast model coupling given measured weather conditions in the 2016/2017, 2016/2017 and 2016/2018 wheat growing seasons in Jessore. Importantly, the calibration procedures described in this

report were conducted for Dinajpur only. This was due to constraints in time and accessing unbroken weather data records; calibration will be undertaken for the remaining two experimental sites in Bangladesh, as well as a range of locations and years in Brazil and presented in the next project report.

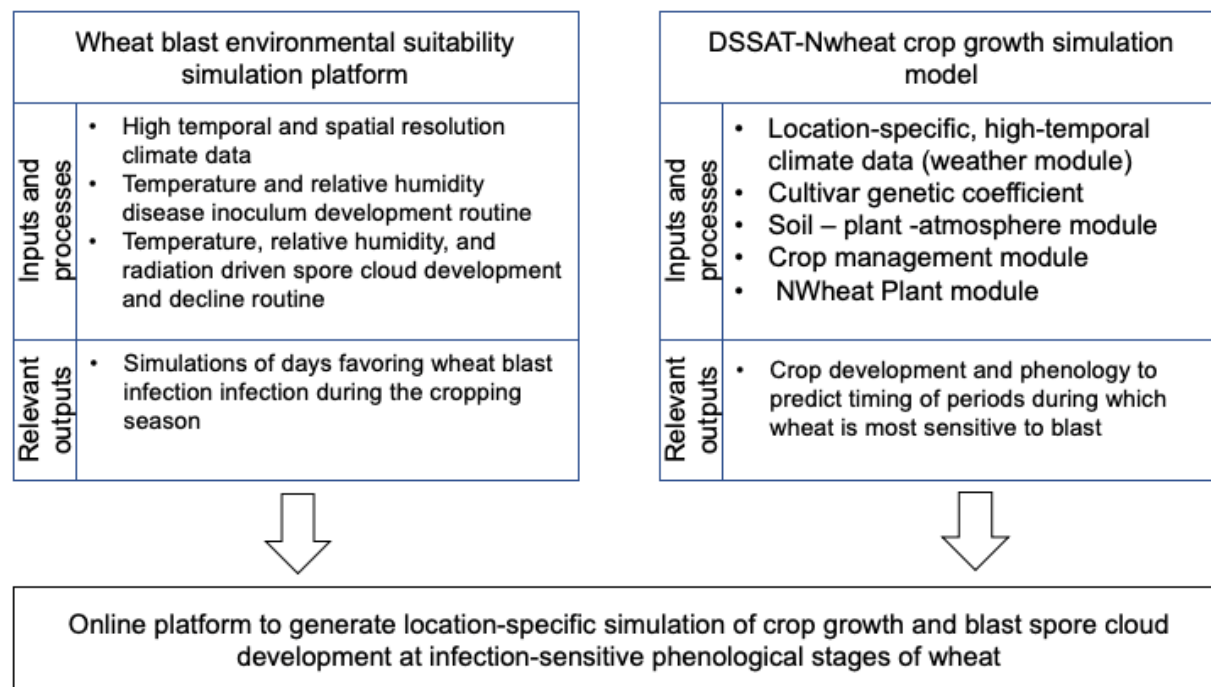


Figure 3. Conceptual diagram of the coupling of wheat blast disease and DSSAT crop growth simulation models to develop an early warning and advisory system to mitigate wheat blast risks.

3. Results and Discussion

3.1 Preliminary DSSAT-Nwheat calibration efforts

Exempting the December 20th sowing date, experimental results from Dinajpur showed a general decline in grain yield as sowing date was delayed from November 25th through January 1 (Figure 4). The inverse relationship between sowing date and in yield has been observed by a number of research groups in South Asia and is associated with terminal heat stress (Farooq et al. 2011; Mondal et al., 2013; Krupnik et al., 2015; Arshad et al. 2018). Heat stress in wheat can cause a reduction in pollen sterility and desiccation of stigma that result in lower seed set and grain filling (Reynolds et al. 2016). Simulated wheat yield and 1,000 grain weight results however showed variable fit with observed data; simulated grain yield was consistently 1-30% lower than observed values. 1,000 grain weight was also between 1–7 g below observed

values, depending on sowing dates. Crop biomass (grain + straw) weights were however consistently underestimated by 15 to 50%.



Figure 4. Observed and simulated values of yield, above ground biomass and grain weight obtained at varying dates of sowing during 2017 in Dinajpur, Bangladesh.

The DSSAT-Nwheat model conversely simulated the number of days after planting (DAP) to anthesis and maturity quite well, with most data falling close to a 1:1 line between simulated and observed data (Figure 5). This indicates only a few days of error. All of these results should however be taken as preliminary and not as definitive. Efforts to further improve model calibration and fit with observed data will be undertaken in the initial months of 2019 for Dinajpur and Jessore.

3.2. Observed weather conditions during years for which blast was simulated

Maximum and minimum temperature patterns in Jessore varied in the years for which simulations were performed. Average temperatures were 24.13/17.64, 25.89/14.21, and 28.31/13.65 °C in 2015/16, 2016/17 and 2017/18 (Figure 6).

This location is characterized by generally high maximum RH, but day-to-day fluctuation that results in a variable minimum RH. Crop water demand was not met by sufficient rainfall in any of these years (hence the need for irrigation). The wheat growing season rainfall totals were 27 mm in 2015/16, 1.1 in 2016/17 and 21.8 mm in 2017/18, the latter associated with a large precipitation event in early December that caused a large increase in RH. Precipitation in the 2015/16 season had four light precipitation events.

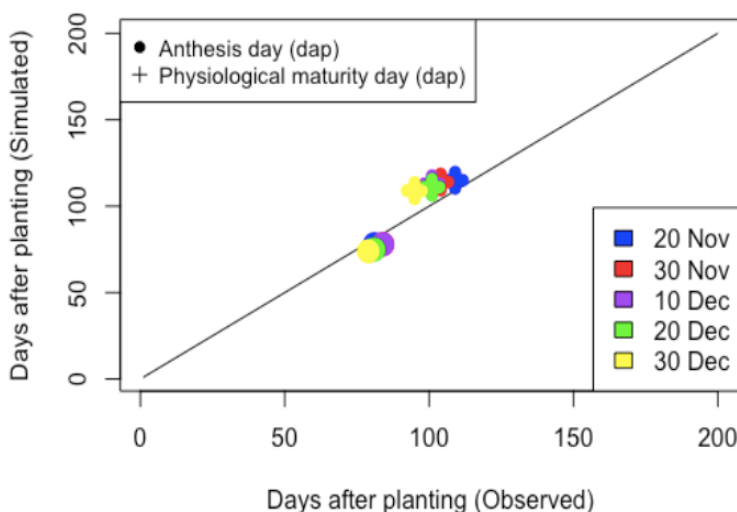


Figure 5. Comparison of observed and simulated time of anthesis and physiological maturity under different sowing dates with the wheat cultivar BARI Gom 26 during 2017 in Dinajpur, Bangladesh.

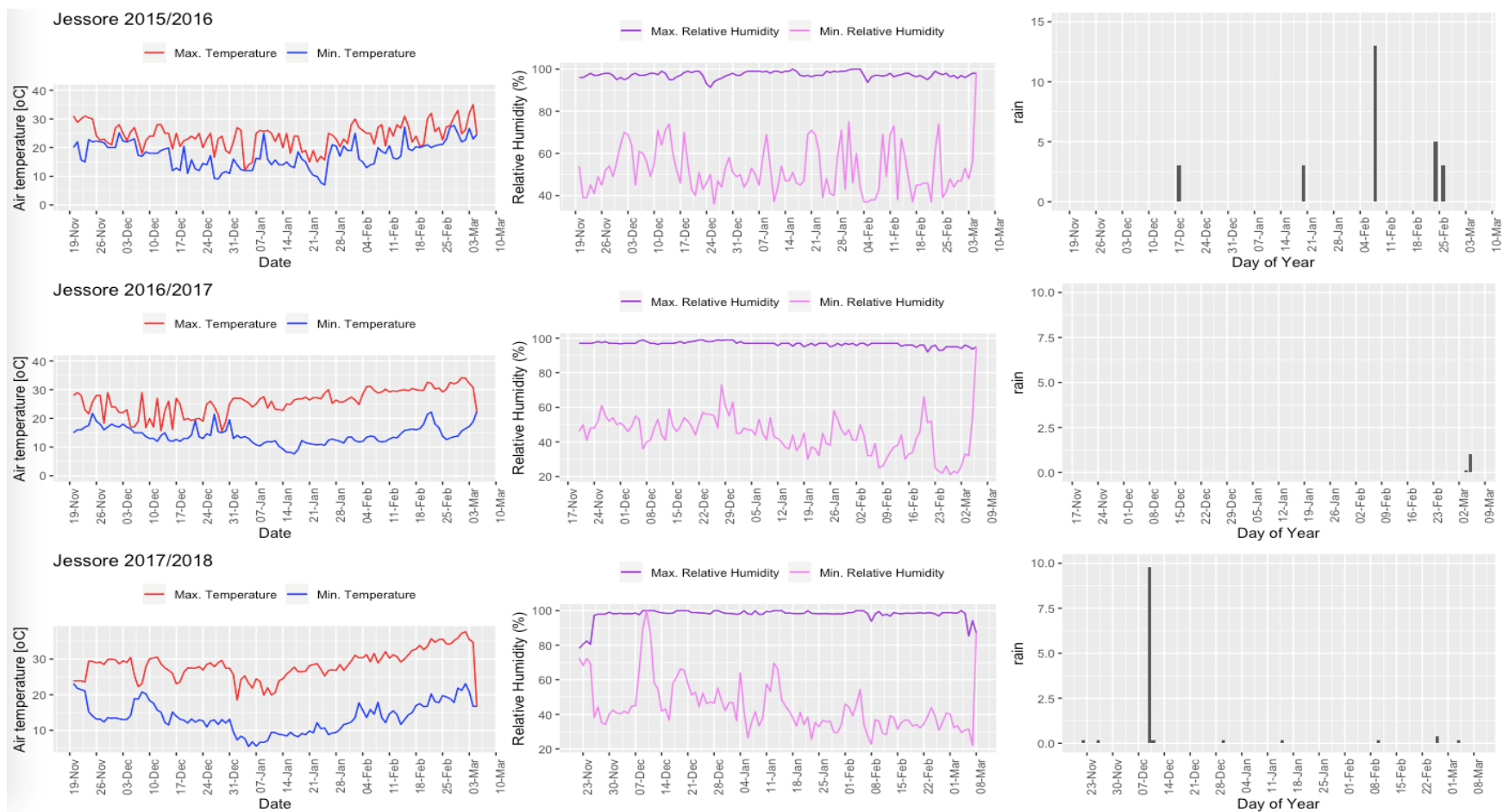


Figure 6. Hourly minimum and maximum temperature, minimum and maximum relative humidity and daily rainfall during 2015/2016, 2016/2017 and 2017/2018 wheat growing seasons, in Jessore, Bangladesh.

3.2. Coupled crop-disease models: Virtual experimentation

An example of the GUI depicting blast disease simulation model outputs for the year 2015/16 in Jessore can be found in Figure 7.



Figure 7. Graphical user interface for the weather-data driven wheat blast disease model with days favoring infection indicated in red. The model integrates temperature and relative humidity to determine spore cloud development and inoculum potential to the wheat crop.



Figure 8. Observed and simulated values of yield, above ground biomass and grain weight obtained at varying dates of sowing, with and without wheat blast, during 2015/16 in Jessore, Bangladesh.

Simulated outputs for virtual experiments conducted using weather data for the 2015/16 wheat season in Jessore showed a grain and biomass yield decline to model runs with progressively later sowing dates without disease (Figure 8). The 2015/16 season was the first season during which large-scale outbreak of wheat blast occurred in Bangladesh (Malaker et al., 2016; Mottaleb et al. 2018). During this season, nearly 15,000 ha were affected (approximately 3.5% of Bangladesh's wheat area) with yield declines reported from 5-51% (Islam et al., 2016). Our preliminary simulation efforts show a reduction in yield of 5-10% in disease. This reduction is relatively light and indicates that further refinements in the model may be necessary to accurately represent yield declines. That said, a 'conservative' range of yield loss from wheat blast was employed by Mottaleb et al. (2018) in an *ex-ante* a climate analogue study that estimated the potential economic impact of wheat blast across South Asia.

When data were simulated for the 2016/17 wheat season on Jessore, both disease-free and diseased crop grain yield showed a decline from the first to last sowing date (Figure

9A). Simulations of blast disease reduced yield only marginally, e.g. from 0-3% relative to disease-free simulations, with a slight increase in disease inflicted yield loss with later sowing dates. Impact on 1,000 grain weight and total biomass was also small. These limited simulated yield loss values for the 2016/17 cropping season are to some extent supported by large-scale field surveillance that was undertaken by the BWMRI, CIMMYT and Cornell University throughout Bangladesh in early 2017. Out of 800 wheat fields surveyed in 25 districts of Bangladesh, only 77 showed symptoms of blast. Those that with blast however all had very low levels of infection and negligent yield or economic impact (data not shown).

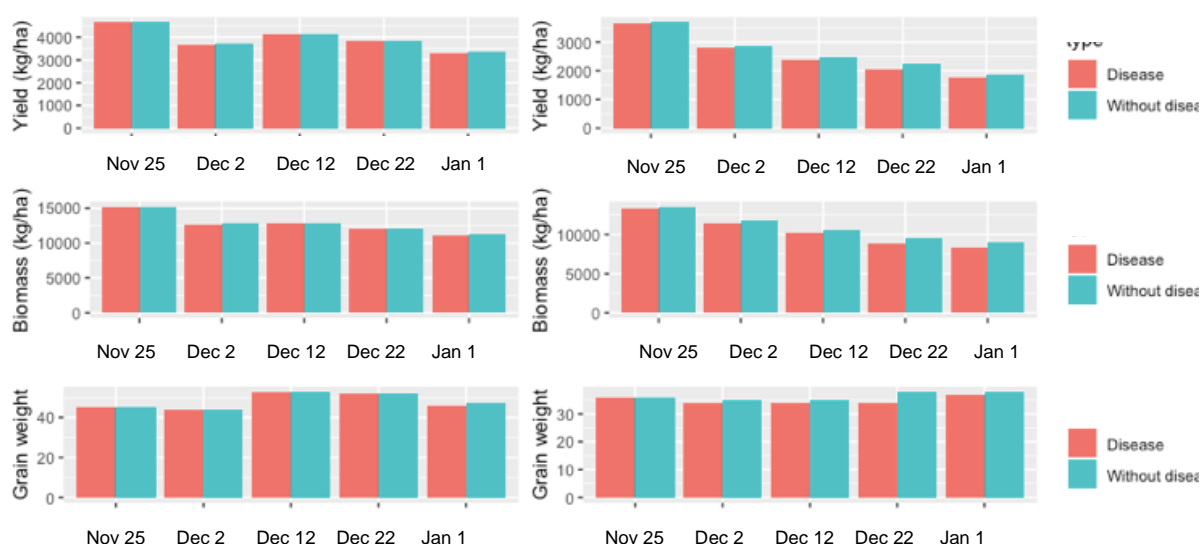


Figure 9. Observed and simulated values of yield, above ground biomass and grain weight obtained at varying dates of sowing, with and without wheat blast, during **(A)** 2016/17 and **(B)** 2017/18 in Jessore, Bangladesh.

Simulations for the Jessore BWMRI station during the 2017/18 wheat growing season showed similar trends, with a gradual decline in simulated wheat yield without blast disease (Figure 9B). Overall, model simulations indicated a potential 3-7% reduction in yield as a result of wheat blast infection (Figure 8B). Little effect of simulated disease was however found for 1,000 grain weight (which ranged in reduction from zero to 2 grams only). Total crop biomass reduction from simulated disease pressure ranged from 6-13%, the latter occurring on the latter two sowing dates. Surveillance from repeat visits to 616 fields across Bangladesh during this season also found very limited infections. Less than 5 ha were found to have infections throughout the country, and those that did registered blast severity <20% (data not shown). While promising, these results are however only very preliminary and should be treated with considerable caution.

4. Implications and future research

This report documents initial efforts to couple a weather-data driven wheat blast development model with the DSSAT-Nwheat model to simulate yield reduction as a function of disease severity in Bangladesh. The original weather-data driven disease model was developed in Brazil and validated for locations where more than a decade of wheat blast observations were available. This model was subsequently transferred to Bangladesh. Disease and crop model coupling will eventually permit model users to explore a range of sowing dates, weather conditions, and cultivar effects on wheat blast infection.

Our very preliminary results showed that the DSSAT-Nwheat model reproduced the general effects of sowing date on wheat phenology and yields. Simulations showed relatively good consistency with the higher yields in earlier sowing dates. Further model calibration efforts undertaken in early 2019 will include direct comparisons of DSSAT-Nwheat simulations to observed experimental data in Jessore. We will also work to relate simulated data to visual estimates of blast incidence and severity at experimental locations in Bangladesh, while also utilizing a broader range of genetic coefficients for other wheat cultivars. The same modeling process will also be used to evaluate the performance of model coupling over more than ten years of data in Brazil.

Outbreak risk maps will be produced over different years using available field trial data to calibrate DSSAT and examine past climatic and crop risks in Bangladesh and Brazil, although we expect that this model can be applied to other countries where data are available. This work will improve upon provisional climatic modeling efforts to estimate the potential effect of wheat blast infection. Mottaleb et al. (2018) for example made use of a climate analogue approach that identifies locations with analogous temperature and precipitation regimes to a location of interest (for example locations where blast infections were observed in Bangladesh in 2016) to estimate the potential geographical area vulnerable to wheat blast. This approach – while initially useful – does not include a biologically approach to estimating when and where blast spores might develop to levels creating outbreaks in wheat. This approach takes work a step further to permit dynamic simulation of potential yield losses from blast given variable weather conditions and climatic regimes.

Although it is still under validation using historical data and field observations, our efforts to estimate the environmentally suitable locations and days during which wheat blast outbreak risks are most severe will incorporate numerical weather model forecast outputs produced by BMD. These will be used to generate time- and location-specific advisories on a five-day forecasted basis. This system will result in the triggering of an automated alert system using the GUI described above. We are also investigating options or delivering blast warnings directly to farmers using interactive voice response (IVR) technologies for at-risk locations prior to potential infection. In contrast to advisories that encourage calendar-based preventative sprays without field scouting, this data-driven

approach can be used to adaptively advise farmers how to take effective and safe preventative action through the intelligent use of fungicides for disease control. This preliminary early warning system is available in beta-format at: <http://dev.sisalert.com.br/shiny/wheatblast/>. In combination with the DSSAT-Nwheat crop model coupling, our future objective is to provide farmers not only with forecast advisories, but also estimates of potential yield loss if they do not take preventative action to control blast. This however requires considerably more additional research that will be discussed in the subsequent report.

Acknowledgments

We thank S. Ishtiaque and R. Sen for supplying the genetic coefficients for BARI Gom 26. This research is supported by a mini-grant from the CGIAR Platform for Big Data, as well as by the USAID funded 'Climate Services for Resilient Development in South Asia' and the 'Training, surveillance, and monitoring to mitigate the threat of wheat blast disease in Bangladesh' projects. The contents and opinions expressed herein are those of the author(s) and do not necessarily reflect the views of USAID, the United States Government or the CGIAR Platform for Big Data and shall not be used for advertising or product endorsement purposes.

References

- Arshad, M., Amjath-Babu, T. S., Aravindakshan, S., Krupnik, T. J., Toussaint, V., Kächele, H. and Müller, K. (2018). Climatic Variability and Thermal Stress in Pakistan's Rice and Wheat Systems: A Stochastic Frontier and Quantile Regression Analysis of Economic Efficiency. *Ecological Indicators* 89: 496-506.
- Batchelor, William & Jones, James & J. Boote, K & Pinnschmidt, Hans. (1993). Extending the use of crop models to study pest damage. *Transactions of the ASAE*. 36. 551-558. <https://doi.org/10.13031/2013.28372>
- Boote, K. J., Jones, J. W., Mishoe, J. W., Berger, R. D. (1983). Coupling pests to growth simulators to predict yield reductions. *Phytopathology*, 73:1581-1587.
- Browne, P. A. and Wilson, S. (2015). A simple method for integrating a complex model into an ensemble data assimilation system using MPI. *Environ. Model. Softw.* 68(C): 122-128.
- Cardoso CDA, Reis EM, Moreira EN (2008) Development of a warning system for wheat blast caused by *Pyricularia grisea*. Summa. *Phytopathol* 34:216–221.
- Calvero SB Jr, Coakley SM, Teng PS (1996) Development of empirical forecasting models for rice blast based on weather factors. *Plant Pathol* 45:667–678
- Cruz, CD and Valent, B (2017) Wheat blast disease: danger on the move. *Trop Plant Pathol*. doi:10.1007/s40858-017-0159-z.

Cruz, C. D., Peterson, G. L., Bockus, W. W., Kankanala, P., Dubcovsky, J., et al. (2016). The 2NS translocation from *Aegilops ventricosa* confers resistance to the Triticum Pathotype of *Magnaporthe oryzae*. *American Society of Agronomy*, 56, 990–1000

Das, S. (2017). Wheat Blast disease enters India from Bangladesh, ICAR official says damage contained. *The Financial Express*, March 6, 2017. Available online: <https://www.financialexpress.com/india-news/wheat-blast-disease-enters-india-from-bangladesh-icar-official-says-damage-contained/576247/>. Verified 3 February 2019.

Deacon, J. (2005) Fungal spores, spore dormancy, and spore dispersal, in fungal biology, 4th edition, Blackwell doi:10.1002/9781118685068.Chapter 10.

Del Ponte, E. M., Fernandes, J. M. C., Pavan, W., & Baethgen, W. E. (2009). A model-based assessment of the impacts of climate variability on fusarium head blight seasonal risk in Southern Brazil. *Journal of Phytopathology*. <https://doi.org/10.1111/j.1439-0434.2009.01559.x>

Duveiller E., He X., and Singh P.K. (2016). Wheat Blast: An Emerging Disease in South America Potentially Threatening Wheat Production. *World Wheat Book, Volume 3. A History of Wheat*. Bonjean A. and van Ginkel M. (Eds.) Pages 1107–1122. Lavoisier, Paris, France.

Farooq, M., Bramley, H., Palta, J. A. and Siddique, K. H. M. (2011). Heat Stress in Wheat during Reproductive and Grain-Filling Phases. *Critical Reviews in Plant Sciences* 30(6): 1-17.

Farman M., Peterson G.L., Chen L., Starnes J.H., Valent B., Bachi M.P., Murdock L., Hershman D.E., Pedley K.F., Fernandes J.M.C., Bavaresco J. (2017). The *Lolium* pathotype of *Magnaporthe oryzae* recovered from a single blasted wheat plant in the United States. *Plant Dis* 101:684–692.

Fernandes, J. M., Lazzaretti, A., Pavan, W., & Tsukahara, R. Y. (2011). Information architecture for crop growth simulation model applications. *International Conference on Information and Communication Technologies*.

Fernandes, J. M. C., Nicolau, M., Pavan, W., Hölbis, C. A., Karrei, M., de Vargas, F. Tsukahara, R. Y. (2017). A weather-based model for predicting early season inoculum build-up and spike infection by the wheat blast pathogen. *Tropical Plant Pathology*. <https://doi.org/10.1007/s40858-017-0164-2>.

Goulart A.C.P., Amabili R.F., Nasser L.C.B., Freitas M.A. (2003). Detection of *Pyricularia grisea* on barley seeds produced under central pivot irrigation in the Brazilian Cerrado. *Fitopatol Bras*. 28:566.

Igarashi S, Utimada C.M., Igarashi L.C., Kazuma A.H., Lopes R.S. (1986). *Pyricularia* em trigo. 1.Ocorrência de *Pyricularia* sp. no estado do Paraná. *Fitopatol Bras*. 11:351–352.

Islam M.T., Croll D., Gladieux P., Soanes D.M., Persoons A., Bhattacharjee P., et al. (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology*. 14: 84. <https://doi.org/10.1186/s12915-016-0309-7> PMID: 27716181.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., and Ritchie, J.T. (2003). The DSSAT cropping system model. *Eur. J. Agron*. 18:235-265.

Kassie, B. T., Asseng, S., Porter, C. H., & Royce, F. S. (2016). Performance of DSSAT-Nwheat across a wide range of current and future growing conditions. *Eur. J. Agron.* <https://doi.org/10.1016/j.eja.2016.08.012>

Krupnik, T. J., Ahmed, Z. U., Timsina, J., Shahjahan, M., Kurishi, A. S. M. A., Miah, A. A., Rahman, B. M. S., Gathala, M. K. and McDonald, A. J. (2015). Forgoing the Fallow in Bangladesh's Stress-Prone Coastal Deltaic Environments: Effect of Sowing Date, Nitrogen, and Genotype on Wheat Yield in Farmers' Fields. *Field Crop. Res.* 170: 7–20.

Lazzaretti A.T, Fernandes J.M.C, Pavan W., Toebe J., Wiest R. (2016). AgroDB - Integration of database management systems with crop models. In: International Environmental Modelling and Software Society (iEMSs) 8th International Congress on Environmental Modelling and Software. Toulouse, France. *iEMSs* p. 194-201.

Malaker, P. K., Barma, N. C. D., Tewari, T. P., Collis, W. J., Duveiller, E., Singh, P. K., Joshi, A. K., Singh, R. P., Braun, H.-J., Peterson, G. L., Pedley, K. F., Farman, M. and Valent, B. (2016). First Report of Wheat Blast Caused by *Magnaporthe oryzae* Pathotype Triticum in Bangladesh. *Plant Disease* <http://dx.doi.org/10.1094/PDIS-05-16-0666-PDN>.

Mondal, S., Singh, R. P., Crossa, J., Huerta-Espino, J., Sharma, I., Chatrath, R., Singh, G. P., Sohu, V. S., Mavi, G. S., Sukuru, V. S. P., Kalappanavar, I. K., Mishra, V. K., Hussain, M., Gautam, N. R., Uddin, J., Barma, N. C. D., Hakim, A. and Joshi, A. K. (2013). Earliness in wheat: A key to adaptation under terminal and continual high temperature stress in South Asia. *Field Crop. Res.* 151: 19-26.

Mottaleb, K. A., Singh, P. K., Sonder, K., Gruseman, G., Tiwari, T. P., Barma, N. C. D., Malaker, P. K., Braun, H. J. and Erenstein, O. (2018). Threat of wheat blast to South Asia's food security: An ex-ante analysis. *PloS one* 13(5): <https://doi.org/10.1371/journal.pone.0197555>.

Reynolds, M. P., Quilligan, E., Aggarwal, P. K., Bansal, K. C., Cavalieri, A. J., Chapman, S. C., Chapotin, S. M., Datta, S. K., Duveiller, E., Gill, K. S., Jagadish, K. S. V., Joshi, A. K., Koehler, A., Kosina, P., Krishnan, S., Lafitte, R., Mahala, R. S., Muthurajan, R., Paterson, A. H., Prasanna, B. M., Rakshit, S., Rosegrant, M. W., Sharma, I., Singh, R. P., Sivasankar, S., Vadez, V., Valluru, R., V., V. P. P. and Yadav, O. P. (2016). An integrated approach to maintaining cereal productivity under climate change. *Glob. Food Sec.* 8: 9-18.

Vales, M., Anzoátegui, T., Huallpa, B., & Cazon, M. I. (2018). Review on resistance to wheat blast disease (*Magnaporthe oryzae* Triticum) from the breeder point-of-view: Use of the experience on resistance to rice blast disease. *Euphytica*, 214(1).

White, J. W. Hunt, L. A. Boote, K. J. Jones, J. W., Koo, J., Kim, S., Porter, C. H., Wilkens, P. W., Hoogenboom, G. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards, *Comput. and Elect. in Ag.* 96:1-12. <https://doi.org/10.1016/j.compag.2013.04.003>.

Appendix 1: The genetic coefficients for the wheat cultivar BARI Gom 26 used in this study

Cultivar	<i>Bari Gom 26</i>
VSEN	1
PPSEN	1.29
P1	520
P5	600
PHINT	120
GRNO	15
MXFIL	2.40
STMMX	3
SLAP1	300
SLAP2	270
TC1P1	2.50
TC1P2	1
DTNP1	5
PLGP1	1400
PLGP2	0.60
P2AF	0.60
P3AF	50

P4AF	3
P5AF	1
P6AF	3
ADLAI	1
ADTIL	1
ADPHO	1
STEMN	0
MXNUP	0.60
MXNCR	40
WFNU	2
PNUPR	450
EXNO3	6.75
MNNO3	0
EXNH4	6.50
MNNH4	0
INGWT	3.50
INGNC	30
FREAR	250

MNNCR	<i>1.230</i>
GPPSS	<i>2</i>
GPPES	<i>5</i>
MXGWT	<i>55</i>
MNRTN	<i>4.50</i>
NOMOB	<i>0.25</i>
RTDP1	<i>1</i>
RTDP2	<i>1</i>

Annex 5

Final report: **Documentation for release of MarkSim base data from WorldClim 2.0**

Peter G. Jones. July 2018 to February 2019

Introduction

From the first records in the 18th century in Europe daily weather has been observed and recorded; it was crucial for the agricultural revolution in England. Researchers have long used these data for agronomic and grazing studies, some of these go back to not long after the records began.

The ethos of data collection for the agricultural world spread to European colonies around the world. These were used to plan and manage plantations. Tea in Kenya and Assam, cacao in western Africa, rubber in the Malay peninsula and cotton in America.

All of these analyses were done on a point basis; if meteorological data were needed then they had to be recorded at the point of use. Early methods of interpolation relied on climate classifications, which could be mapped; Köppen ([1884](#)) is an example. Digital versions of climate data, for example Hijmans *et. al* ([2005](#)) are useful for GIS application where environmental indices can be calculated from the monthly climate data.

Daily data are generally needed for crop modelling, but these are far too voluminous to store in a high-resolution global coverage. This problem is compounded by the need to include realistic variation, both short term and annual, in modelling analyses. One solution to this is to use a weather generator. MarkSim is a daily rainfall generator devised for the tropics in CIAT and ILRI in the late 1990's building on work at CIAT from the previous 15 years; Jones & Thornton, ([1993](#), [1997](#), [1999](#), [2000](#)), Jones *et. al.* ([2002](#)). More recently, see Jones & Thornton ([2013](#)), we have added a capability to simulate future weather sequences from a suite of GCM models, [MarkSimGCM](#) a stand-alone version is available in an EXE as [MarkSim Standalone](#).

The weather generator needs all of the monthly climate values in one record and so the MetGrid file was born. This eventually used the data layers of [WorldClim](#). This work is to update the MarkSim application data with the new data from WorldClim v2.0, Fick & Hijmans ([2017](#)). The outputs are a new set of MetGrid files at 10, 5 and 2.5 arc minutes and 30 arc seconds, and sets of CLI files, which can be used as input to the stand-alone version of MarkSim or as input to other weather generators. Until now, MetGrid files have not been available to the general user; this changes with this version and they will be available for download from CCAFS at CIAT. As they are a highly specialised format, a Fortran module, MetGrid_Handler, has been produced for general distribution to assist their use.

MetGrid

Reason

The basic MetGrid structure was created in the late '80s for various applications in CIAT that used long-term climate data from the CIAT Climate Database. In those days GIS was rudimentary and using single variable coverages for any calculation more complicated than simple overlay was out of the question. The CIAT Climate Database (later to be incorporated into WorldClim) held individual station data. These were interpolated into 10 minute and 5 minute grids, but instead of storing the grids by monthly variates, I devised the MetGrid file so that calculations such as water balance, length of season, climate and land classification and varietal suitability could all be calculated directly, pixel by pixel, using Fortran programs.

Structure

For a start, a MetGrid file is not a grid file. It is a space-conserving file can be used to construct derived climate grids throughout the world. The base data are from WorldClim v2.0 with the exception of the number of rain days that are derived from MarkSim v1.0. MetGrid files come in four resolutions; 10 arc-minutes, 5 arc-minutes, 2.5 arc-minutes and 30 arc-seconds. The big difference from the WorldClim files is that all variates are represented in one record.

The MetGrid files are not raster grids, or, in fact, any sort of grid, and so they have to carry additional data to identify the record. This is simply the Latitude, longitude, and median elevation of the pixel. MetGrid v1.0 used the latitude and longitude of the lower left corner of the pixel, but we found that rounding errors made it difficult to translate easily to normal GIS grids. In version 2 we've decided on the latitude and longitude of the centre of the pixel. The user will now find no problem with translating any MetGrid to a raster image using the functions in the MetGrid_handler module.

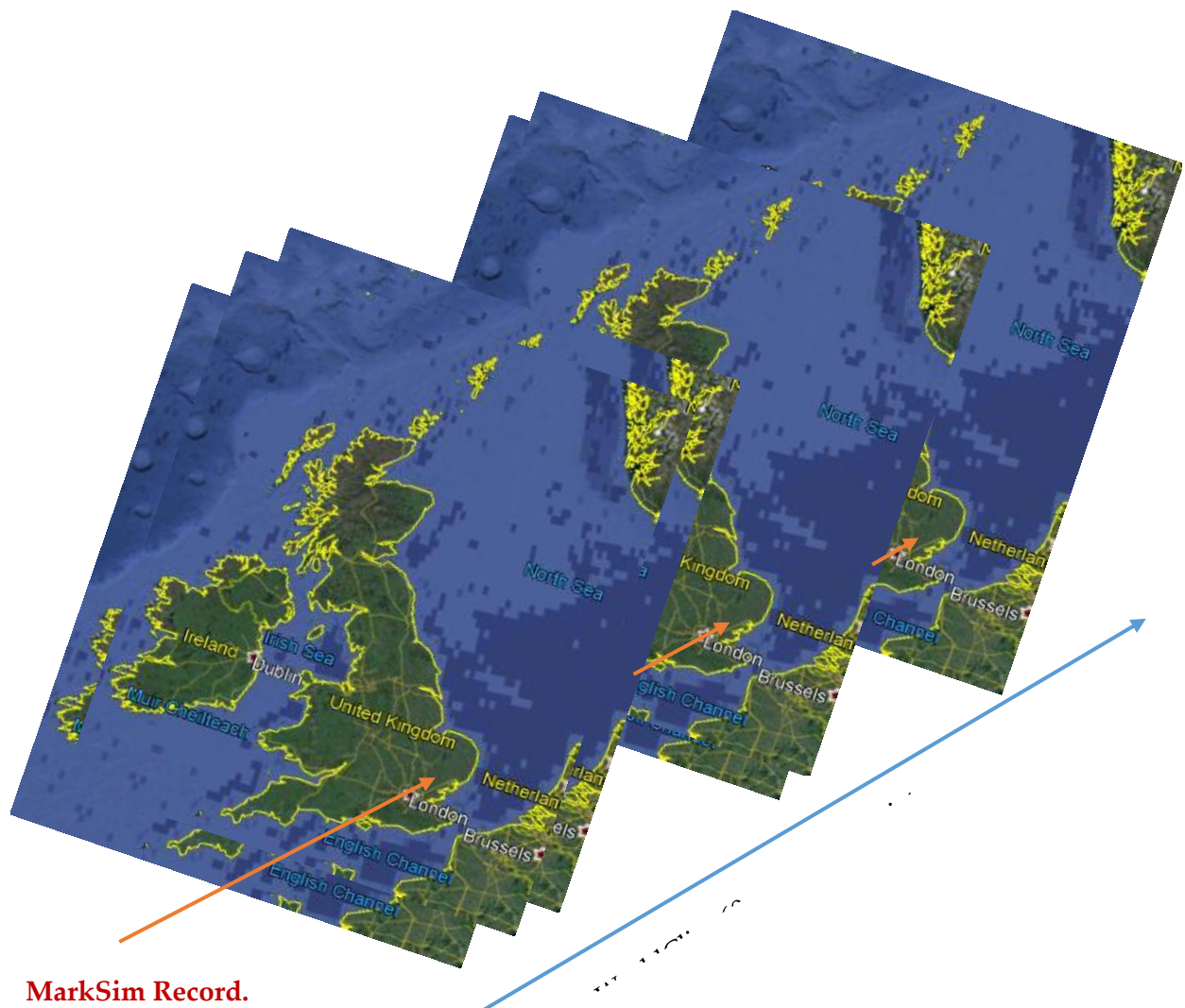


Fig 1, Illustration of the different data structures of MetGrid and WorldClim

Data representation

MetGrid records are stored as binary record in order from North to South in lines of longitude from West to East, but non-land areas are not recorded. There is therefore no simple relationship between a record number in the MetGrid file and a raster map. In order to overcome this the full MetGrid fileset is a group of three files with dedicated file extensions. They should always be copied together as a unit or the full functioning of the MetGrid will be impaired.

They consist of a file header, extension .hdr, which describes the fileset. It is a small ASCII file such as this one for the 10 arc minute dataset. Note that although the contents appear to describe a raster grid dataset, the MetGrid is not; what is being described here is the grid that it represents.

Name World2_10m.mtg


```

Author          P.G.Jones, Jul '18
Cols            2160
Rows            870
Pixels          583580
Max_lat         85.0
Min_lat         -60.0
Max_long        180.0
Min_long        -180.0
Resolution      0.166666667
Index           yes
Index_name      World2_10m.idx
Index_rec_size  57

```

The next file is the MetGrid data file, extension .mtg. This is a binary file of compressed climate data records. It is unreadable without the functions in MetGrid_handler, or ones constructed on similar lines. The record definition is:

```

type metgrid_record                                ! Total record length = 104 bytes
  real*4      : lat, long, phase
  integer*2   : elev
  integer*2   : r (9)                                ! mm
  integer*2   : t (9)                                ! degree C times 30 + 2100 offset
  integer*2   : d (9)                                ! degree C times 100
  integer*2   : s (9)                                ! MJ times 110
  integer*2   : rd (9)                               ! raindays times 120
end type metgrid_record

```

The nine by 2-byte data fields are the twelve monthly values compressed into 12-bit values and scaled as indicated. They are stored as shown below. They do not conform to full word boundaries.

```

012345670123456701234567012345670123456701234567 - bit numbering
1      2      3      4      5      6      7          - bytes
1      2      3      4          - half word
boundaries  aaaaaaaaaabbbbbbbbbbccccccccccddddd etc.      12 bit
words a, b, c, d as packed.

```

The use of 12-bit words in modern computing is unusual, but the type and quality of the data in this case lend themselves to it, considerably reducing record size and hence file size. The extra computation involved can be offset against the reduced seek and read times from the file which will only have to be produced once. The overall record *does* conform to standard 32-bit word boundaries, and so it is read efficiently.

The variates are:

Lat, long	latitude and longitude
Phase	The angle through which the data have been rotated (see the section on data rotation below)
Elev	The modal elevation for the pixel taken from GTOPO30
R	rainfall per day mm
T	mean temperature per month °C

D	mean diurnal temperature range per month °C
S	mean solar radiation per day Mjm^{-2}
Rd	rain days per day. Calculated from the internal functions of MarkSim v1.0

Indexing

The index file (extension .idx) is a binary file that can efficiently relate a latitude and longitude to a single record in the direct access .mtg file without involving multiple reads and searches. It is essentially a run length encoded file by lines of latitude. The run length encoded values are simply presence or absence of a land based pixel and the identity of the target record within the .mtg file can be returned with the absolute minimum of computation. Each .idx is specific to the resolution of the .mtg file and so must never be separated from it.

Rotation

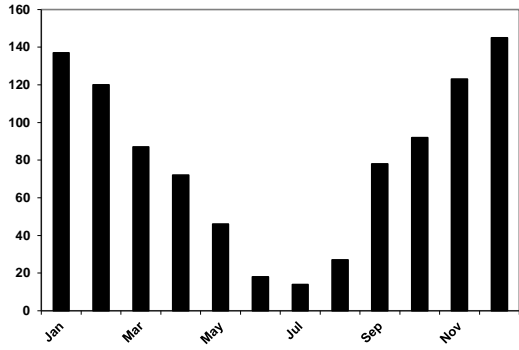
The use of MarkSim and the MetGrid files requires an understanding of climate rotation. The following section was taken mainly from the original manuals of FloraMap, MarkSim and Homologue, where it is crucial in the operation of these applications.

The climatic events that occur through the year, such as summer/winter and start/finish of the rainy season, are of prime importance when comparing one climate with another. Unfortunately, they occur at different dates in many climate types. The most obvious case is where we compare climates between points in the Northern and Southern Hemispheres, but more subtle differences occur in climate event timing throughout the tropics. What we need is a method of eliminating these differences to allow us to make comparisons free of these annual timing effects.

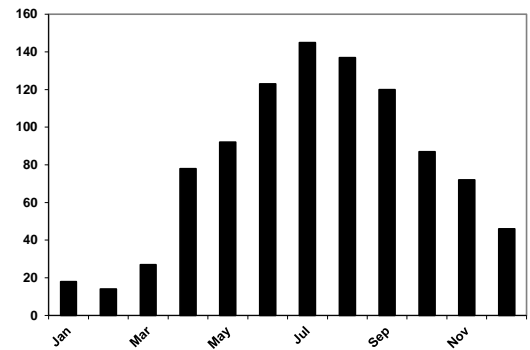
Let us look at two hypothetical climate stations. They are in a typical Mediterranean climate—warm wet winters, hot dry summers. Northville could be somewhere in California, and Southville might be in Chile. The August rainfall in Southville happens in January in Northville (Figure 3.3). If we plot these rainfalls in polar coordinates, we can readily see that to compare them we need to rotate them to a standard time.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Northville	137	120	87	72	46	18	14	27	78	92	123	145
Southville	18	14	27	78	92	123	145	137	120	87	72	46

Table 1 Typical rainfalls for imagined Mediterranean climates



Northville monthly



Southville monthly

Fig 2 Monthly rainfalls for Northville and Southville.

How do we do this automatically? The answer is the 12-point Fourier transform. This is fortunately the simplest of all the possible Fourier transform algorithms. It is highly computationally efficient and fast. It takes the 12 monthly values and converts them to a series of sine and cosine functions. The one used in MarkSim has a modification to make it conserve the monthly total values (Jones, 1987). The equation produced is:

$$r = a_0 + \sum_{i=1}^6 a_i \sin(ix) + b_i(ix) \quad (5)$$

We can write this as a series of frequency vectors, each with an amplitude α_i , and a phase angle, θ_i :

$$\alpha_i = \sqrt{(a_i^2 + b_i^2)} \quad \theta_i = \sin\left(\frac{b_i}{\alpha_i}\right) = \cos\left(\frac{a_i}{\alpha_i}\right) \quad (6)$$

If we subtract the first phase angle from all the other vectors in the set, we have produced a rigid rotation of the vectors; this is the rotation we want, it puts the maximum of the first frequency at a phase angle of zero and places the rest in positions equivalent to their angular separation in the original data. We then use the first phase angle for rainfall to rotate the data for temperature and diurnal temperature range, which we rotate through the same angle.

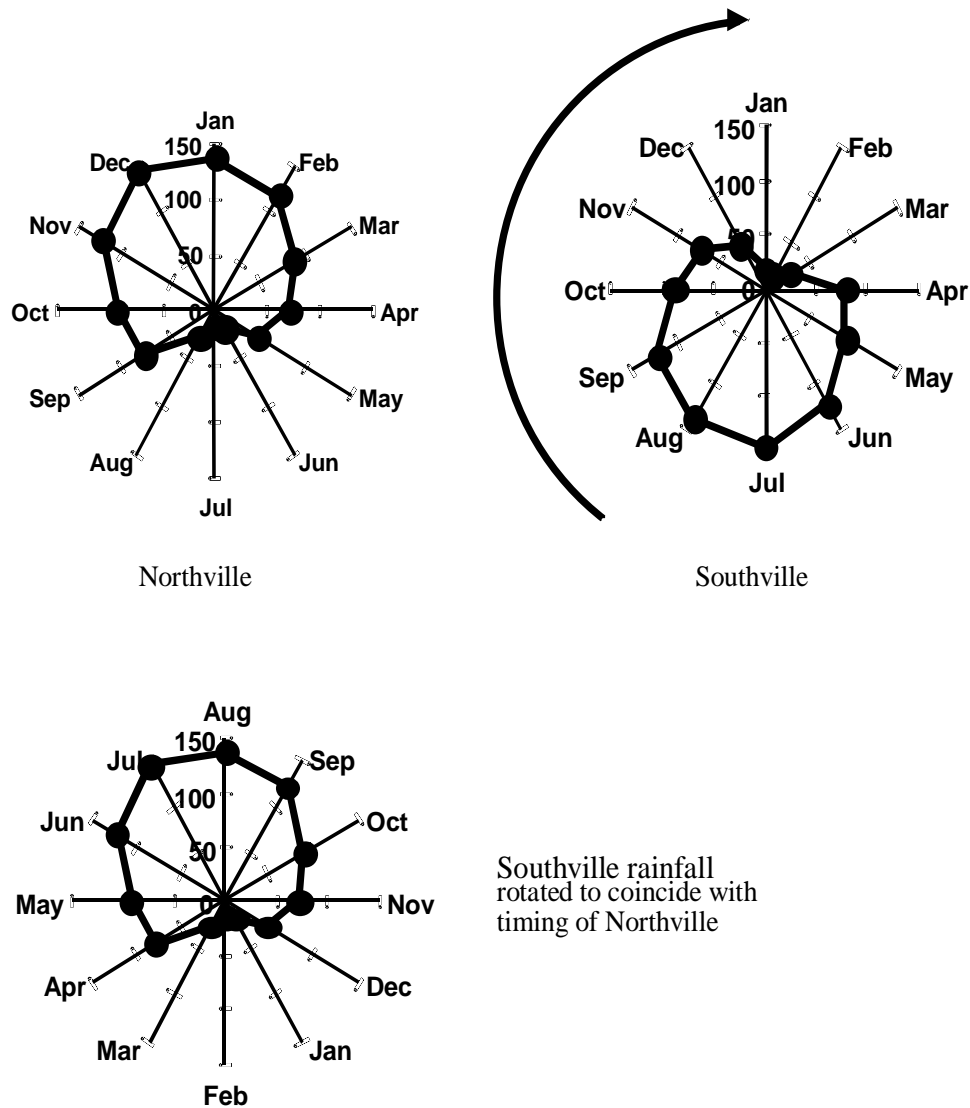


Fig 3 Illustration of the rigid rotation of a MetGrid record

This explanation works well for the tropics. There was a small chance of the procedure going off the rails if the rainfall record did not have a seasonal peak. This was the case in some records from tropical desert regions, in these cases the rotation was ambiguous and sometimes resulted in pixels allocated to the wrong cluster.

The beta release of MarkSim went out with this type of rotation algorithm, as did the first release of FloraMap. When the climate grids of the latter were extended to Europe, the case arose where annual climate pattern was dominated by temperature and not rainfall.

We therefore have the possibility of rotating on rainfall or temperature, but when to decide which is the dominant? We tried many combinations of rules, but unfortunately came to the conclusion that none were acceptable. They all resulted in a hard line across the map at some point where the rotation basis changed. This led to climates that should have been grading imperceptibly from one type to another suddenly jumping at a discontinuity. This would have given the users serious problems when fitting models in these areas.

The best solution found is to use BOTH the rainfall and the temperature in calculating the rotation phase angle. Thus:

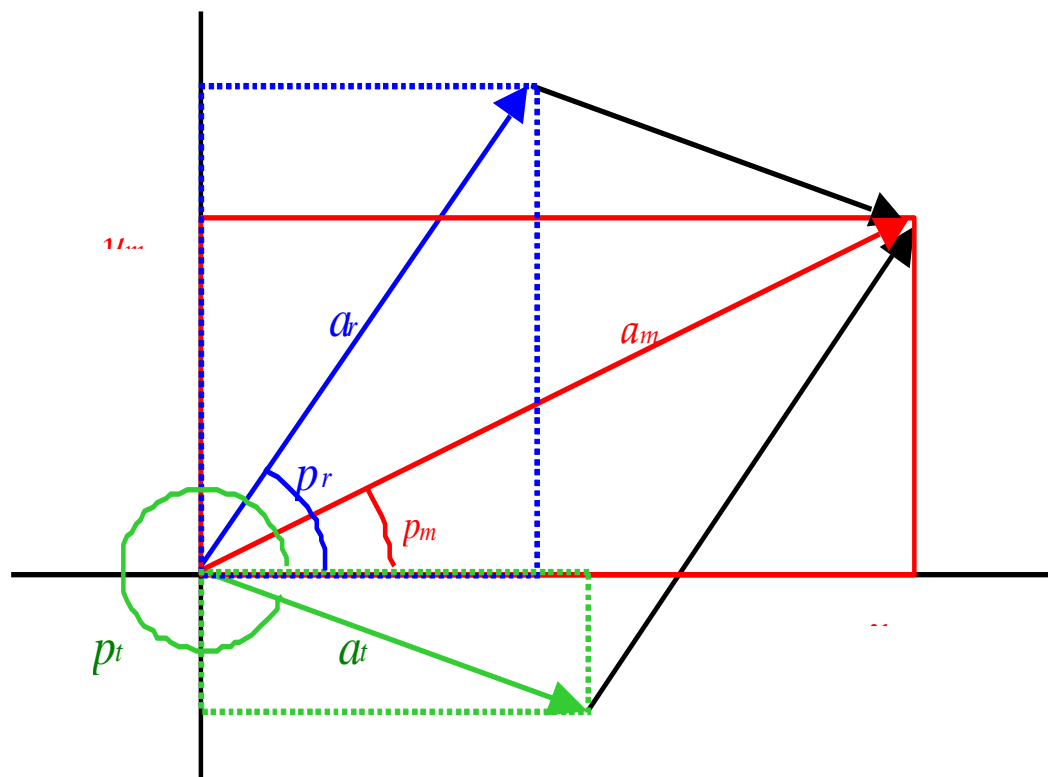


Fig 4 Vector diagram of the first phases of rainfall (a_r) and temperature (a_t) with the resultant vector (a_m)

The resultant phase angle and amplitude are then:

$$y_m = a_r \cos p_r + a_t \cos p_t$$

$$x_m = a_r \sin p_r + a_t \sin p_t$$

$$a_m = \sqrt{y_m^2 + x_m^2}$$

$$p_m = \text{angle} \left(\frac{x_m}{a_m}, \frac{y_m}{a_m} \right)$$

Unfortunately, this does not completely solve the problem of fitting a model to climates with different weather determinants. However, the vast majority of climates in the world are either:

- (1) Rainfall determined where temperature is not an important seasonal effect (large areas of the tropics and subtropics);
- (2) Temperature determined where rainfall is even throughout the year (most of the rest of the tropics and some temperate climates); or
- (3) Rainfall and temperature determined when the two variates are highly correlated (summer rains - most of the rest of the world).

The Odd Man Out is:

- (4) Winter rains and hot dry summers (almost only Mediterranean climates).

Luckily, the Mediterranean climates are at moderately high latitudes and we can afford to have the rotation dominated by temperature without losing generality in the rotations and comparisons. We therefore need to increase the weighting for the temperature vector smoothly as we approach the Mediterranean climates (in order to avoid a sudden swing). I found the following weightings to work well: **p**=rainfall mm and

t = temperature $\times 2 \times \text{abs}(\text{latitude})$.

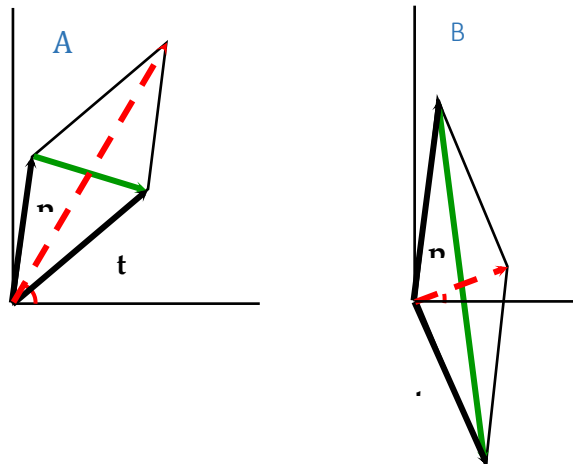


Fig 5 Illustration of concordant and discordant weighted rotation vectors

In A the two vectors of **p** and **t** (the bold black lines) are pointing in the same direction, in B they are pointing in disparate directions. The dashed red line shows the sum of the vectors and the green line shows their difference. From these I can calculate an index of reliability of the rotation as $\arctan(\text{sum/difference})$

The index marked OK at 0.79 indicates (see Fig 7) an angle of 90 degrees between the two vectors if they are of equal length that denotes a reasonably stable rotation angle. It also corresponds to the case where one vector dominates the phase angle, which is also acceptable.

The perfect confidence is found in the subtropics there the two vectors point in same direction.

The equatorial regions and some Mediterranean regions have vectors pointing to different degrees in opposite direction. Luckily, none of these areas reaches the level where the phase angle is completely indefinite. The highest index noted is about 1.1, which is equivalent to an angle of about 120 degrees between two vectors of the same length.

While this is somewhat indeterminate, there is still enough purchase to get a unique phase angle for rotation

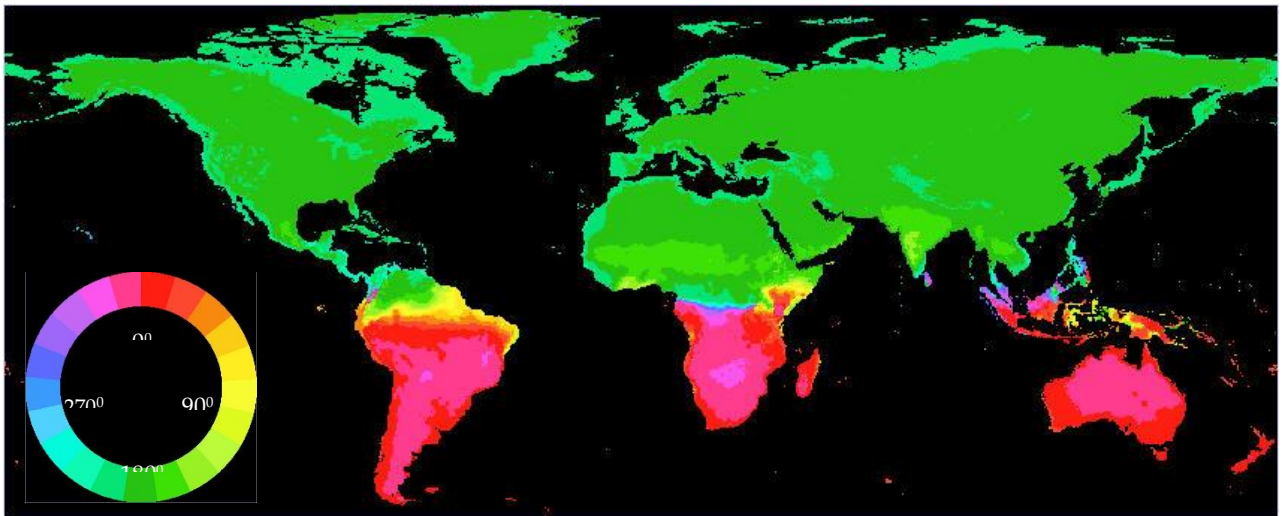


Fig 6 The rotation angle of world climates.

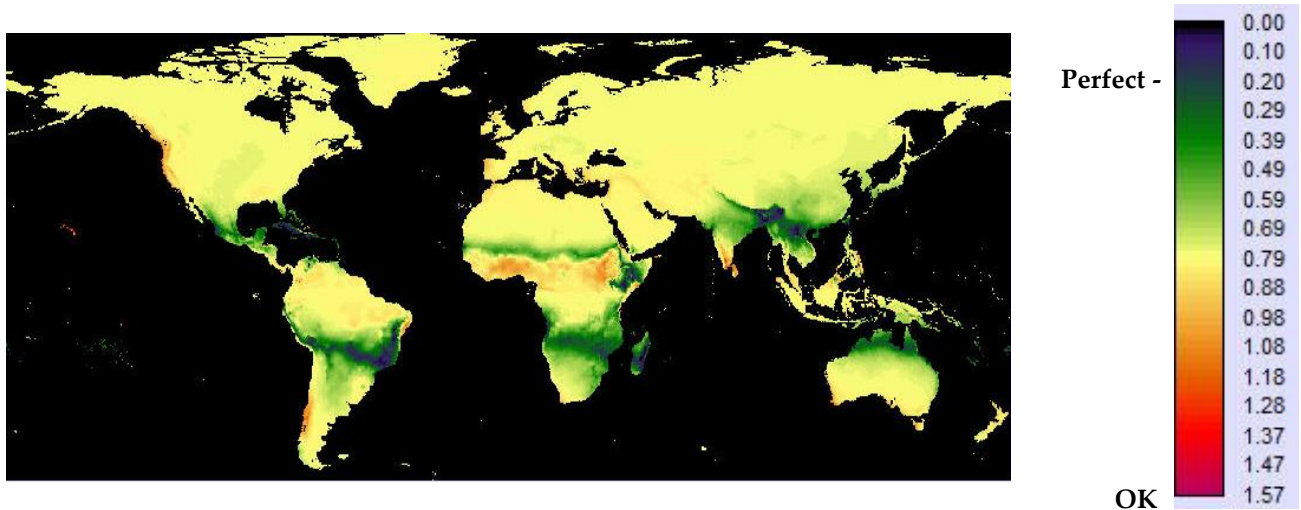


Fig 7 Confidence in the rotation angle of the world climates.

MetGrid_Handler module

Accessing MetGrid files from Fortran

A Fortran module contains data structure definitions that specify defined type variables that allow the grouping of various types of data into structured data units, and sets of functions or subroutines that use these units or operate on them. The module is invoked from within a Fortran program with the *use* command thus: use MetGrid_Handler

Module Description

Variable structures

The structure most likely to be interpreted by the user, in that the internal variables will be used in a Fortran program is the `climate_structure`. Most of the others are specific to the internal workings of the module and so are mostly only minimally described. This structure normally only exists during a run unit of the program.

```

type climate_structure
  real*4      ::      lat,    &      ! decimal degrees latitude
                                long,  &      ! decimal degrees longitude
                                phase, &      ! radians rotation angle
                                confidence ! unused when constructed from MetGrid record
  real*4      ::      rain(12) ! mm per day
  real*4      ::      temp(12)  ! mean temperature degrees centigrade
  real*4      ::      diurn(12) ! diurnal temperature range degrees centigrade

  real*4      ::      srad(12)   ! solar radiation Mj m-2 day-1

```



```

        real*4           ::      rainedays(12)    ! dimensionless 0 to 1
        integer*4        ::      elev            ! metres above sea level
        logical*1         ::      rotated         ! climate structure may be rotated or not
end type climate_structure

```

```

type cli_record          A structure to hold the data for a CLI record
type Fourier            The Fourier coefficients of a 12 monthly climate array
type metgrid_record     The compressed form of the climate_structure
type index_element      A 3 byte element of the index file holding a count of land or sea pixels
type metgrid_hdr        A structure to hold the information contained in a MetGrid header file.
This                    structure is held in the public variable 'h' once it is loaded.

```

```

type metgrid_hdr
  character*20           :: name
  character*50           :: author
  integer*4              :: cols, rows, pixels
  real*4                 :: Max_lat, Min_lat, Max_long, Min_long, Resolution
  logical*4              :: index ! 'yes' or 'no ' in the file - translates
                           ! logical on input
to
  character*20           :: Index_name
  integer*4              :: Index_rec_size ! max(n * type(index_element))
  logical*4              :: loaded = .false.
end type metgrid_hdr

```

```

type pane_descriptor    Used for the production of compressed panes of CLI files but a written file
                        version is present in the RAR compressed pane of CLI files and can be
                        inspected to verify the contents of the pane

```

Public variables

Some variables are used within the module, but are defined as public as they may be of use in a program using the module. These include halfpi, pi and twopi as real*4, eof, year, month and day as integer*4 and month_code as an array of 12 three byte character representations of the calendar month. If the MetGrid header has been read then it is held in the variable h. Lastly, two logical*4 variables header_loaded and index_loaded indicate if the respective loading has been done.

The index arrays are defined as private and are only accessed within the module; module routines to access MetGrid records use the loaded index arrays.

Function descriptions

As with the variable structures, not all module functions are necessarily called by the end user. Some have to be included because they are called by end-user functions; others are there because they are of use in the administration of the files.

Grid functions

These four function rely on having the MetGrid header loaded in public variable h

- G01 integer*4 function grid_col (long)**
 returns the column of the encompassing grid for the MetGrid
 parameter longitude real*4
- G02 integer*4 function grid_row (lat)**
 returns the row of the encompassing grid for the MetGrid
 parameters latitude real*4
- G03 real*4 function grid_lat (row)**
 returns the latitude of the encompassing grid for the MetGrid
 parameter row integer*4
- G04 real*4 function grid_long (col)**
 returns the longitude of the encompassing grid for the MetGrid
 parameter column integer*4

Calendar functions

- C01 pure real*4 function month_days (month, year)**
 returns the days of the month
 parameters month integer*4
 year integer*4, optional
 If year is present, the correct day number for the year is given, if year is absent then the value for February is 28.25
- C02 pure integer*4 function rotate_month (month, phase)**
 returns the month number of any month of a rotated record
 parameters month integer*4, unrotated calendar month number
 phase real*4, rotational phase angle
- C03 pure logical function leap (year)**
 returns .true. for a leap year

Input functions

- I01 type (metgrid_record) function find_met_record (lat, long, unit, error)**
 uses the index to read the MetGrid record at lat, long
 parameters lat, long real*4, latitude & longitude
 unit integer*4, logical unit on which the MetGrid is opened
 error logical*4, optional
 error returns all errors are hard stop errors if error not specified
 1. Fault - Index indicates water at this point
 2. Index not loaded at find_met_record
 3. Error reading from MetGrid file on unit
 error 1. returns .true. if error parameter specified
- I02 type (metgrid_record) function nearest_met_record (lat, long, max_distance, unit)**
 uses the index and a spiral search to find and read the nearest record to lat, long. The purpose of this function is to ensure finding a record near a coastline or small island where

the given coordinates may slightly differ from those in the coverage used to create the MetGrid.

parameters	lat, long	real*4. latitude & longitude
	max_distance	real*4, kilometre limit to the search pattern
	unit	integer*4, logical unit on which the MetGrid is opened
error reports	all errors are hard stop	
	1, Index not loaded in Nearest_met_record	
	2, Please reconsider your Maximum distance limits. The function 'nearest_met_record' is designed to trim the MetGrid file record retrieval to other GIS coverages that may have marginal differences.	
	At lat latitude, your search area would be nnn pixels wide.	
	3, Max distance exceeded in nearest_met_record	
	Error 2 avoids exceptionally large searches	

I03 type (cli_record) function read_cli (filename)

reads a CLI file into a Cli_record; this is mainly used for debugging.

parameter	filename	character*(*), the full path and filename of the CLI file
error reports	all errors are hard stop. see module text for error messages.	

I04 type (metgrid_hdr) function read_metgrid_header (path, version)

reads the MetGrid header into the public variable h.

parameters	path	character*(*), the path to the directory of the MetGrid files
	version	integer*4, this is the number of pixels in a degree of latitude or longitude. This can take the values 6, 12, 24 or 120 and determines the resolution of the MetGrid
error reports	all errors are hard stop; see module text for error messages, if the MetGrid files are installed correctly in the directory in path there should be none	

I05 subroutine load_metgrid_index (path, version)

reads the MetGrid index into private arrays in the module. Parameters and errors as in I04 if the MetGrid header has not been read, it will do this for you.

I06 subroutine open_metgrid_files (path, version, unit)

opens the relevant MetGrid file on logical unit 'unit'. If the header and index have not been loaded, it will do this for you.

parameters	path	character*(*), the path to the directory of the MetGrid files
	version	integer*4, this is the number of pixels in a degree of latitude or longitude. This can take the values 6, 12, 24 or 120 and determines the resolution of the MetGrid
	unit	integer*4, logical unit for the MetGrid file. Do not use the unit 9999 here or in any of your programming as it is used internally by the module.

I07 subroutine close_metgrid_files (unit)

This MUST be used if MetGrid files are to be opened subsequently in the program. It is not enough to close (unit) as the index arrays must be deallocated.

Output functions

These routines are principally used for development and debugging, none have any use in actual analysis programs; they are largely self-descriptive.

- O01** `subroutine print_climate(c)`
- O02** `subroutine print_metgrid_record (m)`
- O03** `subroutine write_climate(c, unit)`
- O04** `subroutine write_cli (cli, version, path)`
- O05** `subroutine write_pane_descriptor (p, path)`
- O06** `subroutine print_header`

Basic structure operations

These are important routine for the end user as they control the movement of data between MetGrid format and the more accessible climate_structure. There are no error states in any of these functions as it is assumed that they always operate on valid structures

- B01** `type (climate_structure) function make_climate_structure (m)`
 this function takes a MetGrid_record and converts it into a climate_structure. A MetGrid file is always rotated so the resulting climate structure will be rotated. The basic operation of this routine is to unpack the data and present them in a form easily used by a Fortran program
- B02** `type (climate_structure) function null_climate ()`
 this function creates an empty climate structure. It is not possible to zero out a climate_structure by a simple replacement operation without this function
- B03** `type (cli_record) function make_CLI_from_climate_structure (c)`
 takes all of the information from a climate structure and calculates the variates for a CLI. the CLI.code is always 'MARK' and the CLI.title and CLI.source are blank and need to be set by the user after calling this routine.
- B04** `type (climate_structure) function rotate_climate (c, phase)`
 care must be taken when using this function as it can do one of three things depending on the call and the state of the climate_structure.
 parameters c type (climate_structure), input will not be changed unless specified as the function target.
 phase real*4, optional. a phase angle radians, may be any magnitude but will be used modulo twopi.
 1) If phase is specified the structure is rotated through the angle, 'phase' irrespective of whether the structure is rotated or not.
 2) if phase is not specified and the structure is rotated, it will be rotated to zero phase angle i.e. true calendar time.

3) if phase is not specified and the structure is not rotated, then a new, natural, phase angle will be calculated for the structure and it will be rotated to this angle.

B05 type (metgrid_record) function compress_met_record (c)

This function transforms a climate_structure into a MetGrid_record. This is used in creation of the MetGrids, but can be used as a handy way to save climate_structures by writing them to a binary file, the record length should be 104 bytes.

B06 type (metgrid_record) function null_metgrid_record ()

As for null_climate, this is a way of creating an empty metgrid_record.

Rotation operations

None of these routines need concern the end user as they are all subsumed into rotate_climate. However, I will briefly describe them so that the end user may gain an insight into what is happening. They all have to be present in the module because they are interdependent and used on the working of rotate_climate.

R01 type (climate_structure) function new_rotate (c)

This function creates a new climate_structure such that the rotation angle is the definitive angle that rotates the structure to match similar climates. It cannot be used if the structure is already rotated

R02 subroutine neg_correct (v)

This routine correct for small errors that are sometimes introduced during rotation. These are notices on variates that cannot go negative, such as rain and raindays. They are an artefact of the rotation process, but are invariably small and can be eliminated
parameter v 12 valued real*4 array, declared as input and output

R03 subroutine rotate (v, phase)

Rotate a 12 valued variate array through a phase angle. Although there may be times when this routine is appropriate, it is dangerous as the array is divorced from any structure and no internal knowledge of the state of previous rotation is known.

parameters v 12 values real*4 variate array, declared as input and output
phase real*4, rotation phase angle

R04 subroutine decode (q, v)

This, and R05, are applications of the 12 point Fourier transform described in Hamming (1973). However there is a difference, the routines here have been recalculated to fit the curve, not through the mid-point of the month, but to the monthly integral of the data. In other words, the daily values of each month sum to the total for that month, and therefore monthly means are conserved

parameters q real*4, 13-valued array of Fourier coefficients
v real*4, 12 values variate array

R05 subroutine encode (v, q)

The inverse of R04

R06 subroutine freq1 (q, f)

Convert a 13-valued array of Fourier coefficients into polar form.

parameters	q	real*4, 13-valued array of Fourier coefficients
	f	type (Fourier) variate containing mean, and six phase and amplitudes

angles

R07 subroutine frqinv1 (f, q)

The inverse of **R06**

R08 real*4 function angle (sin, cos)

Returns the angle in radians from 0 to twopi related to the parameters sin and cos

parameters	sin, cos	real*4 sine and cosine of the required angle
------------	----------	--

R09 subroutine frrota (q, phase)

Rotate an array of Fourier coefficients through a phase angle

parameters	q	real*4, 13-valued array of Fourier coefficients
	phase	real*4, rotation phase angle

Miscellaneous operations

M01 subroutine pack12 (v, n, p)

Pack a 12 valued data array into 18 bytes. See [packing](#)

parameters	v	real*4, 12 values variate array
	n	integer*4, index number of data within Climate_structure
		1 = rain per day
		2 = temperature
		3 = diurnal temperature range
		4 = solar radiation
		5 = probability of rain (raindays per day)
		integer*2, 9 valued packed array

M02 subroutine unpack12 (p, n, v)

The inverse of M01

M03 type (pane_descriptor) function pane (row, col, version)

Constructs the variate pane_descriptor.

parameters	row	integer*4, grid row and column pertaining to the MetGrid
	col	integer*4, grid row and column pertaining to the MetGrid
	version	integer*4, version number describing the resolution

M04 integer*4 function version_index (version)

For some instances, it is useful to have an index of the resolution. This routine returns the index value (1,2,3,4) for the resolutions (6,12,24,120)

Examples

Example 1

Text

```
! PURPOSE: recover the climate data for Dolgellau,
!          and print the august rainfall
!*****

program example_1

use MetGrid_Handler
implicit none

type (metgrid_record)      :: m
type (climate_structure)   :: c
logical*4                  :: error

call open_metgrid_files ('c:\worldclim_2\MetGrids\' ,120,1)
m = find_met_record(52.737,-3.883, 1, error)
if(error) then
    print *, 'coordinates must be wrong'
    stop
end if
```

```

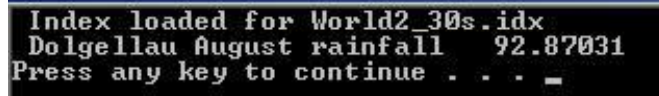
c = make_climate_structure (m)                                !5
c = rotate_climate (c)                                       !6
print *, 'Dolgellau August rainfall', c.rain(8)*month_days(8) !7

stop

end program example_1

```

Output



```

Index loaded for World2_30s.idx
Dolgellau August rainfall  92.87031
Press any key to continue . . .

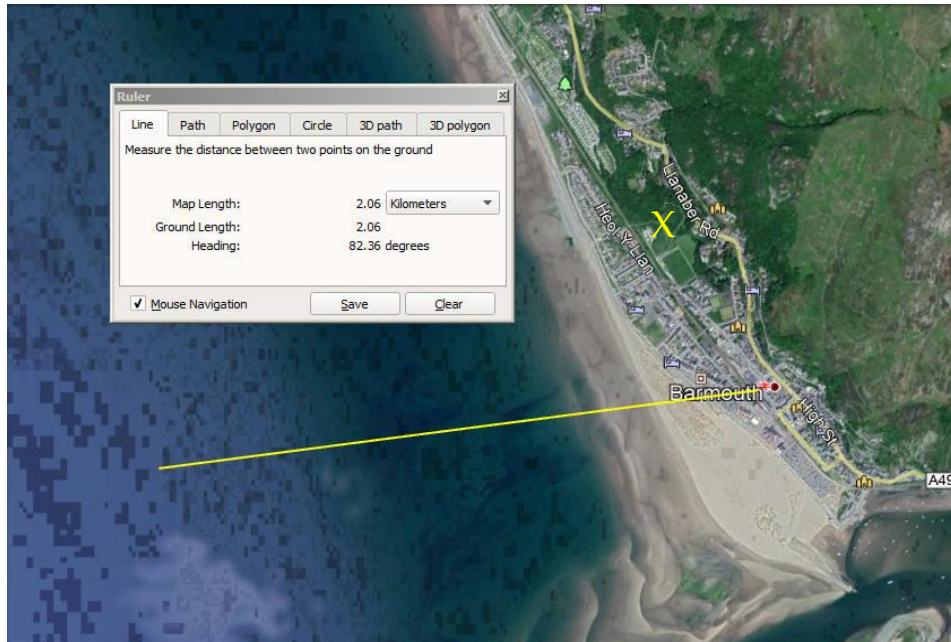
```

Fig 7 Example 1 output

Comments

- !1 Definition of the structure to hold the [MetGrid record](#)
- !2 Definition of the structure to hold the [Climate structure](#)
- !3 Opens the [MetGrid files](#) in the directory with path 'c:\worldclim_2\MetGrids\'. The version is 120 pixels per degree, that is the 30 arc second version, the MetGrid file itself is opened as binary, direct access on unit 1.
- !4 Use the index to [find the record](#) at latitude 52.737, longitude -3.883, Dolgellau, actually the Unicorn Inn. read from the MetGrid file on unit 1 and return an error if the position turns out to be in the sea.
- !5 [Make](#) the climate structure from the MetGrid record
- !6 The MetGrid is always [rotated](#). This is because the majority of the uses for a MetGrid file require the data in the rotated form. This line [rotates the record](#) to normal calendar form so that August is actually month 8.
- !7 Rainfall is [stored](#) as rain per day. This has the advantage of being independent of month length, however [this is available](#) from the module.

[Example 2](#)



Find Abermaw (Barmouth) with coordinates from an unreliable map.

Fig 8 Google Earth map of Barmouth showing a possible 2.1km error from an unreliable map

Text

```
! PURPOSE: finding the nearest MetGrid record
!           when coordinates may be uncertain
! *****
```

```
program example_2
  use metgrid_handler
  implicit none

  type (metgrid_record)      :: m
  type (climate_structure)   :: c

  call open_metgrid_files('c:\worldclim_2\MetGrids\' , 120, 1)

  m = nearest_met_record(52.72, -4.087, 10.0, 1)      !1
  c = make_climate_structure(m)
  c = rotate_climate(c)
  call print_climate(c)

  stop

end program example_2
```

Output

```

C:\Worldclim_2\work\worldclim\example_2\Debug\example_2.exe
Index loaded for World2_30s.idx
Lat 52.712 Long -4.029 Elev 43
Phase 0.000 rotated F
Rain 3.05 2.51 2.62 1.96 1.79 2.25 2.17 2.73 2.92 3.44 3.82 3.47
Temp 5.09 4.92 6.39 8.12 11.18 13.74 15.57 15.54 13.46 11.03 7.78 6.12
Diurn 5.32 5.58 5.82 7.08 8.03 7.47 7.13 7.08 6.83 6.17 5.62 5.38
Srad 2.32 4.51 8.29 13.39 17.53 18.66 17.40 14.42 10.19 5.82 2.86 1.80
raindays 0.398 0.342 0.377 0.312 0.301 0.348 0.314 0.405 0.387 0.403 0.434 0.408
Press any key to continue . . . _

```

Fig 9 Output from Example 2.

Note that the routine has found the nearest land based pixel which is actually about 750 metres NNE of the town centre, see the yellow X on Fig 8.

Comments

- !1 Find and read the nearest land pixel to 52.72 North, 4.087 East. Limit the search to 10 km and use the MetGrid open on unit 1.

Example 3

This is an example of how having all of the monthly variables together in one record can be handy for calculations. That the index is pretty useless is irrelevant, still, it brings up the hilly terrain quite nicely. I apologise for slipping in the image_processing module which is not available on this site. Phil Thornton and I use it extensively to produce Idrisi Images from Fortran programs but I've never got round to documenting it. If any would like a copy please contact me at p.jones@cgiar.org

Text

```

! PURPOSE: To map the rainfall/temperature index for North Wales
!(this is not a recognised, or useful index, but serves to illustrate a point)
!*****

program Example_3
use MetGrid_Handler
use image_processing ! Note this module is available from P.Jones@cgiar.org
                     ! but is not yet=documented

implicit none
real*4, parameter      :: NE(2)=[53.5, -5.0], SW(2)=[52.3, -3.0] ! bounds
type (idrisi_doc)       :: d
type (metgrid_record)   :: m
type (climate_structure) :: c

real*4, allocatable     :: im(:, :)
integer*4               :: rows(2), cols(2), row, col
logical*4               :: error

call open_metgrid_files('c:\worldclim_2\MetGrids\ ', 120, 1)
rows(1) = grid_row(NE(1))
rows(2) = grid_row(SW(1))
cols(1) = grid_col(NE(2))
cols(2) = grid_col(SW(2))
d.data_type = 'real'

!1
!2

```

```

d.file_type      = 'binary'
d.cols           = cols(2)-cols(1) + 1
d.rows           = rows(2)-rows(1) + 1
d.ref_system     = 'latlong'
d.ref_units      = 'degrees'
d.unit_distance = 1.0
d.min_X          = NE(2)
d.max_X          = SW(2)
d.min_Y          = SW(1)
d.max_Y          = NE(1)
d.resolution     = h.resolution
d.flag_value     = -9999.0
d.title          = '  North Wales rainfall/temperature index'
allocate (im(cols(1):cols(2),rows(1):rows(2)))
im = -9999.0

do row = rows(1),rows(2)
  do col = cols(1),cols(2)
    m = find_met_record(grid_lat(row),grid_long(col), 1, error)
    if(error) cycle !3
    c = make_climate_structure(m) !4
    im(col, row) = sum(c.rain/c.temp)
  end do
end do

d.min_value      = minval(im, im.ne.-9999.0)
d.max_value      = maxval(im)
d.display_min    = d.min_value
d.display_max    = d.max_value

open (10,file='c:\worldclim_2\data\N_Wales_example.rst',form='binary')
write (10) im
close (10)
open (10, file='c:\worldclim_2\data\N_Wales_example.rdc')
call write_idrasi_rdc(d, 10)

stop

end program Example_3

```

Output

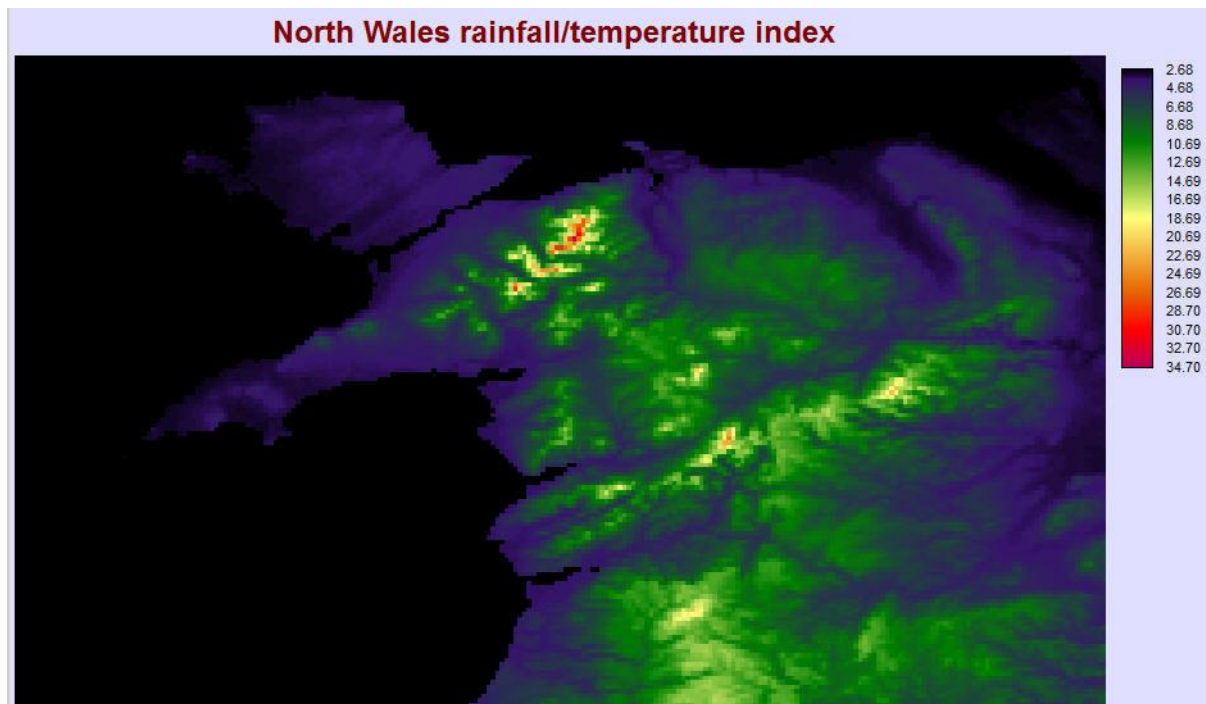


Fig 10 North Wales annual sum of monthly rainfall/temperature index at 30 arc seconds

Comments

- !1 [Grid functions](#) are basic to using the MetGrid file in a grid oriented way. They take all relevant information from the [MetGrid header](#), which must be loaded.
- !2 These are statements relevant to the production of the Idrisi image, you could use any other format for the production of the map image. They are indented one extra tab to indicate that they are not basic to the calculation.
- !3 Error return is used here to skip all other processing when a pixel is all at sea.
- !4 There is no need to [rotate the structure](#) in this case because we are only interested in the total of the monthly indices. Of course the rotation angle is unlikely to fall such that monthly values are exactly conserved. I will leave it to the interested reader to produce the alternative map with the structure rotated back to standard calendar to see what difference this would make. Hint! You don't need to produce two images, merely do two calculation on each pixel and subtract one from the other. The resultant image will be the difference of the two.

CLI files

Reason

These are basic input files for the DSSAT weather generators and are used as input to the executable version of MarkSim ([MarkSim Standalone](#)) which allows researchers without detailed Fortran knowledge to run multiple geographic points for interface with DSSAT. This is particularly useful for mapping crop responses over geographic areas. We chose the DSSAT format because it is widely known and understood and can be adjusted to local climate data where these are available and when they don't match the baseline WorldClim 2.0

Structure

CLI file format

The format for the CLI files closely follows the initial definition part of the standard DSSAT CLI file format (DSSAT, [2003](#), [2017](#)). The following is an example from the MetGrid at 2.5 arc minutes

```
*CLIMATE : R01921C04801
```

```
@ INSI      LAT      LONG  ELEV  TAV   AMP  SRAY  TMXY  TMNY  RAIY
  clim    20.021    20.021   546   25.1   7.8  22.7  32.6  17.5    5
@START  DURN  ANGA  ANGB REFHT WNDHT SOURCE
   1     30   0.25   0.50 -99.0 -99.0 MetGrid v2.0 2.5 minute
@ GSST  GSDU
   0     365
```

```
*MONTHLY AVERAGES
```

```
@  MTH  SAMN  XAMN  NAMN  RTOT  RNUM  SHMN  AMTH  BMTH
   1  18.1  23.5   8.4   0.0   0.1 -99.0   0.2   0.5
   2  20.6  26.1  10.1   0.0   0.1 -99.0   0.2   0.5
   3  22.9  30.4  14.0   0.0   0.1 -99.0   0.2   0.5
   4  23.9  35.3  18.6   0.1   0.1 -99.0   0.2   0.5
   5  25.6  38.1  22.4   1.3   0.1 -99.0   0.2   0.5
   6  25.9  39.1  24.1   0.0   0.1 -99.0   0.2   0.5
   7  26.7  37.7  24.3   1.1   0.1 -99.0   0.2   0.5
   8  26.2  37.0  24.1   3.0   0.1 -99.0   0.2   0.5
   9  24.0  36.9  22.6   0.3   0.1 -99.0   0.2   0.5
  10  21.7  34.0  18.6   0.1   0.1 -99.0   0.2   0.5
  11  19.4  28.7  13.4   0.0   0.1 -99.0   0.2   0.5
  12  17.4  24.6   9.5   0.0   0.1 -99.0   0.2   0.5
```

CLI pane format

CLI pane descriptor

The CLI pane descriptor is both a file that can be found in the compressed copy of each pane and a Fortran type descriptor for the information contained therein. The file looks like this:

```
Name          N020E020
Directory      North_20\East_020\
```

Lat, Row	20.000	1921
Long, Column	20.000	4801
Pixels/degree	24	
Total pixels	14400	

Directory Structure

As can be seen above, the pane descriptor has a directory definition. This is more for the analysts implementing the application than for the end user, but it can be used to structure the directories for the panes if the user so wishes. The directories define 40 degree squares to hold copies of the compressed panes of CLI files.

What constitutes a pane?

The panes are made of CLI files from a geographic square of 120 pixels on the side. This means that the pane sizes are 20 degrees of latitude and longitude for the 10 arc minute resolution and 10 and 5 degrees for the 5 arc minute and 2.5 arc minute ones respectively. The maximum number of files in a pane is therefore 14400, but this is only found in panes that do not include a coastline.

The pane name is constructed from the coordinates, in degrees, of the south western corner of the pane, it is given in the pane descriptor file and is used as the name of the compressed structure, i.e. N020E020.rar. The CLI files are named by row and column, thus: R01921C04801.CLI. Note that the row and column numbers are those of the virtual grid of the MetGrid file and not row and column within the pane. This is to ensure that all CLI files have unique names, however the names will not be unique between resolutions so it is important to keep the three sets separate if you are using more than one. The SOURCE field with the CLI file will define the resolution, but this would involve reading the file to determine its resolution.

Compression

The CLI files and the pane descriptor file are compressed using WinRAR with the -s option to benefit from compressing multiple files with the same structure. This yields a compression to half the size of that obtained from compressing the files without this option. The panes can be uncompressed with WinRAR or WinZip.

User Access

The compressed panes have a maximum size of about 3 Mb and can be downloaded individually from a world map interface at http:*****,
(this description to be completed once the files and interface have been installed)